Final Technical Report

Interface Properties of Wide Bandgap Semiconductor Structures

R. F. Davis, S. Bedair*, J. Bernholc[†], R. J. Nemanich[†], and Z. Sitar
c/o Materials Science and Engineering Department,
*Electrical and Computer Engineering Department
and [†]Department of Physics
North Carolina State University
Campus Box 7907
Raleigh, NC 27695-7907

DTIC QUALITY INSPECTED 4

December, 1997

19980326 012

Table of Contents

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December, 1997	Final Technical 6/1/97–12/31/97

4. TITLE AND SUBTITLE

Interface Properties of Wide Bandgap Semiconductor Structures

5. FUNDING NUMBERS

uri41114a-01
1114SS
N00179
N66005
4B855

6. AUTHOR(S)

R. F. Davis, S. Bedair, J. Bernholc, R. J. Nemanich, and Z. Sitar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

North Carolina State University
Hillsborough Street
Raleigh, NC 27695

8. PERFORMING ORGANIZATION REPORT NUMBER

N00014-92-J-1477

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Sponsoring: ONR, 800 N. Quincy, Arlington, VA 22217-5660
Monitoring: Administrative Contracting Officer, Office of Naval Research
Regional Office Atlanta, 101 Marietta Tower, Suite 2805
101 Marietta Street
Atlanta, GA 30323-6145

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)   Page 1 of 2

It is shown that surface adsorbates affect the electron affinity of diamond; hydrogen leads to a negative electron affinity; and oxygen termination and adsorbate free surfaces exhibit a positive electron affinity. Thin metallic layers can also lead to an effective negative electron affinity of diamond and some surfaces are stable to ambient air exposure. The energy distribution of field emitted electrons from intrinsic cubic boron nitride (c-BN) coated Mo tips was analyzed to determine the origin of the emitted electrons. The emission current at a given voltage increased an order of magnitude with the c-BN coated emitters relative to bare emitters. The energy of field emitted electrons from the c-BN coated emitters was linearly dependent upon the applied voltage. Extrapolation of data from c-BN coated emitters to a flat-band condition revealed that the electrons were emitted from the conduction band minimum of the c-BN coating at the c-BN/vacuum interface. The theory of interfaces and surfaces in III-Nitrides and a consistent approach to the calculation of surface energy valid for all crystal systems are presented. The surface energies of GaN and SiC were determined. The doping properties of substitution C, Si and Ge impurities in GaN and AlN were also investigated using quantum molecular dynamics. Silicon and Ge are shallow donors in GaN; Ge is a deep donor in AlN. Substitutional C on the cation site is a shallow donor in GaN, but a deep one in AlN; C on the N site is a relatively shallow acceptor in both materials. Thin films of AlN and GaN were deposited on a(6H)-SiC(0001) wafers using metalorganic vapor phase epitaxy and H_2 and N_2 diluents. A computational fluid dynamic model of the deposition process was used to analyze the film growth conditions for both diluents. Low temperature (12 K) photoluminescence of the GaN films grown in N_2 had peak intensities and full widths at half maximum of ~7 meV which were equal to or better than those films grown in H_2. Cross-sectional and plan-view transmission electron microscopy of GaN films grown in both diluents showed similar microstructures with a typical dislocation density of $10^9/cm^2$. Hall measurements of n-type (Si doped) GaN grown in N_2 revealed Hall mobilities equivalent to those films grown in H_2. Acceptor-type behavior of Mg-doped GaN grown in N_2 was repeatably obtained without post-growth annealing. Secondary ion mass spectrometry revealed equivalent levels of H in Mg-doped GaN films grown in

14. SUBJECT TERMS

diamond, cubic boron nitride, aluminum nitride, gallium nitride, indium nitride, III-Nitrides, silicon carbide, adsorbates, electron affinity, hydrogen, negative electron affinity, emission current, theory, surface energies, doping, acceptor, donor, C, Si, Ge, MOVPE, growth, dislocation density, transmission electron microscopy, secondary ion mass spectrometry, H,

15. NUMBER OF PAGES

142

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLAS	UNCLAS	UNCLAS	SAR

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December, 1997	Final Technical 6/1/97–12/31/97

4. TITLE AND SUBTITLE

Interface Properties of Wide Bandgap Semiconductor Structures

6. AUTHOR(S)

R. F. Davis, S. Bedair, J. Bernholc, R. J. Nemanich, and Z. Sitar

5. FUNDING NUMBERS

uri41114a-01
1114SS
N00179
N66005
4B855

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

North Carolina State University
Hillsborough Street
Raleigh, NC 27695

8. PERFORMING ORGANIZATION REPORT NUMBER

N00014-92-J-1477

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Sponsoring: ONR, 800 N. Quincy, Arlington, VA 22217-5660
Monitoring: Administrative Contracting Officer, Office of Naval Research
Regional Office Atlanta, 101 Marietta Tower, Suite 2805
101 Marietta Street
Atlanta, GA 30323-6145

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* Page 2 of 2

both diluents. Exposure of selected areas of particular GaN samples with He-Cd laser light (3.8 eV) revealed a persistent and marked decrease in the near band edge photoluminescence (PL) intensity emitted from these areas. This ability to modulate PL emission intensity at individual points in these materials can be exploited as a method for optical data storage. A means of erasing information stored using this effect has also been investigated using lower energy (~2 eV) photons. The microstructures and PL spectra were determined for $In_{x_2}Ga_{1-x}N$ films ($x \le$ ~0.23) grown on substrates of a(6H)-SiC(0001) wafer/AlN buffer layer/GaN heterostructures by low pressure MOVPE at 780°C using nitrogen as the diluent and carrier gas and V/III ratios as low as 2,420. Indium droplets were not observed. The maximum InN content achievable at 45 Torr was ~13%. Increasing the deposition pressure to 90 Torr increased the maximum InN content to ~23%. Room temperature and 12 K PL spectra of the films revealed single-feature, near band edge (NBE) emission with increasing full width at half maximum (FWHM) values with increasing In fraction. The PL NBE FWHM for an $In_{0.23}Ga_{0.77}N$ film at 12 K was 103 meV. InGaN films, 0.3 to 0.5 mm thick, in the 0-50% InN composition range have been analyzed by q-2q x-ray diffraction, transmission electron microscopy and selected area diffraction for the occurrence of phase separation. As-grown films with up to 21% InN were single phase. Films with 28% InN and higher showed a spinodally decomposed microstructure, as confirmed by TEM and extra spots in SAD patterns that corresponded to multiphase InGaN. Metal-insulator-semiconductor (MIS) capacitors were fabricated on n-type GaN (0001) films using thermally grown Ga_2O_3, remote plasma enhanced chemical vapor deposited (RPECVD) SiO_2, and MBE-derived AlN as the gate insulator and Al as the gate electrode. Nitrogen-free polycrystalline films of Ga_2O_3 were grown on the GaN. Capacitance-voltage measurements of capacitors fabricated from this oxide showed distinct depletion and accumulation regions with significant leakage. The AlN and SiO_2 capacitors demonstrated better electrical characteristics than the Ga_2O_3 because of lower leakage. The RPECVD SiO_2/GaN heterostructures, in particular, showed good agreement with curves calculated for an ideal oxide; they contained a small amount of hysteresis.

14. SUBJECT TERMS

photoluminescence, optical data storage, $In_xGa_{1-x}N$, spinodal decomposition, microstructure, metal-insulator-semiconductor capacitors, gate insulator, Ga_2O_3, remote plasma enhanced chemical vapor deposition, capacitance-voltage, depletion, accumulation, hysteresis

15. NUMBER OF PAGES

142

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLAS	UNCLAS	UNCLAS	SAR

I. Introduction

Heteroepitaxy is the growth of a crystal (or a film) on a foreign crystalline substrate that determines its orientation. Such oriented growth requires that lattice planes in both materials have similar structure. In general, an epitaxial relationship is probable whenever the orientation of the substrate and overgrowth produces an interface with a highly coincident atomic structure having low interfacial energy relative to a random arrangement.

During the past decade, nonequilibrium techniques have been developed for the growth of epitaxial semiconductors, superconductors, insulators and metals which have led to new classes of artificially structured materials. In many cases, the films were deposited on substrates having a different chemistry from that of the film, and heteroepitaxy was achieved. Moreover, layered structures with a periodicity of a few atomic layers have also been produced by the sequential heteroepitaxial deposition of a film of one type on another. Metastable structures can be generated which possess important properties not present in equilibrium systems. A consideration of the materials under consideration for next generation electronic and optoelectronic devices, e.g., the III-V nitrides show that only a few of them can currently be grown in bulk, single crystal form having a cross-sectional area of >3 cm^2. Thus other, commercially available substrates must be used. This introduces a new set of challenges for the successful growth of device quality films which are not present in homoepitaxial growth and which must be surmounted if these materials are to be utilized in device structures.

In addition to providing structures which do not exist in nature, applications of advanced heteroepitaxial techniques permit the growth of extremely high quality heterostructures involving semiconductors, metals, and insulators. These heterostructures offer the opportunity to study relationships between the atomic structure and the electrical properties of both the film itself and the interface between the two dissimilar materials. They also allow the study of epitaxial growth between materials exhibiting very different types (ionic, covalent, or metallic) of bonding.

While the potential of heteroepitaxial deposition has been demonstrated, significant advances in theoretical understanding, experimental growth and control of this growth, and characterization are required to exploit the capabilities of this process route. It is particularly important to understand and control the principal processes which control heteroepitaxy at the atomic level. It is this type of research, as well as the chemistry of dry etching via laser and plasma processing, which forms the basis of the research in this grant.

The materials of concern in this report are classified as wide band gap semiconductors and include diamond, SiC and the III-V nitrides of Al, Ga, and In and their alloys. The extremes in electronic and thermal properties of diamond and SiC allow the types and numbers of current and conceivable applications of these materials to be substantial. However, a principal driving force for the interest in the III–V nitrides is their potential for solid-state optoelectronic devices

for light emission and detection from the visible through the far ultraviolet range of the spectrum.

The principal objectives of the research program are the determination of (1) electron emission from CVD diamond cold cathodes and cubic BN coated molybdenum emitters, (2) theoretical studies of III-Nitride surfaces and the doping of these materials, (3) growth and characterization of AlN and GaN thin films, and (4) growth and characterization of $In_xGa_{1-x}N$ thin films, and selection, growth and characterization of gate insulators on GaN surfaces for use in high power field effect transistors.

This is the eleventh and final bi-annual report since the initiation of the project. The following sections introduce each topic, detail the experimental approaches, report the latest results and provide a discussion and conclusion for each subject. Each major section is self-contained with its own figures, tables and references.

Electron Emission from Diamond and Cubic Boron Nitride

II. Electron Emission from CVD Diamond Cold Cathodes

Peter K. Baumann [a] and Robert J. Nemanich [b]

[a] Materials Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

[b] Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

15.1 Abstract

This review focuses on electron emission from diamond. The electron affinity is a measure of the barrier to emission of electrons from the conduction band into vacuum. The effect of surface properties on the electron affinity of semiconductors is described in terms of the surface dipole. In particular, it is shown that surface adsorbates can affect the electron affinity of diamond, and hydrogen leads to a negative electron affinity while oxygen termination and adsorbate free surfaces exhibit a positive electron affinity. Thin metallic layers can also lead to an effective negative electron affinity of diamond and some surfaces are stable to ambient air exposure. Field emission is a more complicated process that appears to be due to electrons in the valence band or from defects in the gap. Yet, the field emission is dependent on the surface termination.

15.2 Introduction

Electron beams are fundamental to many electronic applications ranging from cathode ray tube (CRT) displays to microwave or power amplifiers. While hot cathodes are suitable for many applications, the development of cold cathodes could lead to improved performance in many existing applications and, more importantly, to new technologies including vacuum micro-electronics, flat panel display technologies and new types of microwave amplifiers. A new approach being considered for these applications is to employ semiconducting materials in which electrons in the conduction band can be emitted directly into vacuum without overcoming an energy barrier. This property of the semiconductor has been termed a negative electron affinity (NEA).

The first evidence of this possibility was reported for diamond. Himpsel et al. [1] and Pate [2] reported a high quantum efficiency for photoelectron emission from (111) surfaces of natural diamond samples. It was concluded that these surfaces did indeed exhibit a true negative electron affinity, and hydrogen termination was found to induce induce this effect on the (111) surface. These studies first highlighted the potential of diamond as a cold cathode source material.

As diamond film deposition techniques have been developed since the first photoemission studies, there has been substantial interest in studying the potential of diamond as the emitting material in electron emission structures and devices.

Cold electron emission from metals by means of high electric fields (i.e. field emission) has been studied for many years.[3] And it has been known that low workfunction metals emit electrons more readily that metals with a higher workfunction. The emission process for a semiconductor is more complicated than for a semiconductor. Considering a semiconductor electron emission structure, electrons must be supplied to the semiconductor and then extracted with by an electric field at the surface. The field emission

process then involves, injection of electrons from an electrical contact into the semiconductor, transport of the electrons through the bulk to the emitting surface and finally the emission from the surface into vacuum. The initial studies of Himpsel et al. [1] and Pate [2] demonstrated that unlike metal surfaces, the emitting surface will not limit the emission for hydrogen terminated (111) natural diamond surfaces.

In this report recent studies pertaining to electron emission based on diamond are reviewed. The relation of the electron affinity to the surface properties is presented. The determination of the electron affinity by means of photoemission and secondary electron emission is described. Results for diamond surfaces are reported. Since field emission will be required for most device applications, representative measurements are presented. The complexity of the measurements are discussed.

15.3 Electron Affinity and Negative Electron Affinity

15.3.1 Definition of Electron Affinity and Negative Electron Affinity

The electron affinity of a semiconductor is defined as the energy difference between the vacuum level and the conduction band minimum both extrapolated to the surface. This corresponds to the energy necessary to excite an electron from the conduction band minimum to the vacuum. We also note that the vacuum level is the energy of an electron at rest in vacuum. The free electron model essentially describes the band structure of the vacuum. It is probably worth noting that the electron affinity of the semiconductor is essentially the heterojunction band offset between the semiconductor and vacuum.

In general, the electron affinity is independent of the position of the Fermi level. We make this point since the work function of semiconductors are sometimes quoted, but in general, the work function may be different for p- and n-type doping. For most

semiconducting materials the vacuum level lies above the conduction band minimum, and by convention this has been defined as the electron affinity. For wide bandgap semiconductors like diamond, the conduction band minimum will be near to the vacuum level. And in fact in some instances the vacuum level is below the conduction band minimum. This case has been termed a negative electron affinity or simply NEA. In this case electrons present in the conduction band have sufficient energy to overcome the workfunction of the surface and can be emitted into vacuum.

There are many ways to view the energetics of a semiconductor surface, but the following is helpful in understanding some effects to be described here. The electron affinity of a semiconductor may be determined by (1) the properties of the material itself as well as (2) the surface termination including adsorbates, reconstructions and steps.[4] A schematic illustrating of these effects is shown in Fig. 1.[5] The atomic levels essentially reference the ionization energy of the atom to the levels that broaden into the valence band. Since the atomic levels are more or less intrinsic to a material, they cannot be changed (but alloys may provide a degree of variation). At the surface of any material effects can lead to the formation of a surface dipole. {We note that it is difficult to define the surface dipole exactly, but in much of what follows we will examine how various processes increase or decrease the value of the surface dipole.} For a simple free electron metal, the surface dipole would arise from the quantum mechanical extension of the electron wavefunctions into the vacuum beyond the surface. This also results in a positive layer due to the loss of this electron density. The combination of the electron density away from the surface with the positive charge layer results in a surface dipole that effectively holds electrons in the material.

However, the surface dipole can be influenced substantially by surface adsorbates, surface reconstructions and steps on the surface. These effects may either increase or decrease the electron affinity of the semiconductor. Ignoring the specifics of bonding and

7

charge distribution, a molecular adsorbate that pulls electrons from the surface towards the adsorbate will increase the electron affinity while an adsorbate that contributes electrons to the material will result in a lower electron affinity.

To illustrate the magnitude of this effect consider a hydrogen passivated surface. Let's assume that the average nuclear and electronic charges are point charges separated by 0.5Å. Then for a surface density of $1 \times 10^{15} cm^{-2}$ a surface dipole induced energy shift of about 9 eV can be calculated. (Certainly complete charge transfer is never a reasonable possibility, but this simple calculation demonstrates the significance of the surface dipole.) Full quantum mechanical calculations have addressed the properties of the H terminated diamond surface, and they will be mentioned below. Since the effect of the surface dipole is so large, it is basically impossible to determine if a material is "intrinsically NEA." Thus the surface termination is critical in describing the electron affinity (or NEA) properties of a material.

15.3.2 Techniques

While UV-photoemission measurements first detected the high quantum efficiency of electron emission, the technique of UV-Photoemission spectroscopy (UPS) is a very sensitive method to determine whether a surface exhibits a NEA or to measure the positive electron affinity.[1,2] In this technique, the incident light excites electrons from the valence band into states in the conduction band. Some of these electrons quasi - thermalize to the conduction band minimum. For NEA surfaces these secondary electrons may be emitted into vacuum and are detected as a sharp feature at the low energy end of photoemission spectra. A careful measurement of the width of the photoemission spectrum can be used to determine if the low energy emission occurs from the conduction band minimum. The width represents the energy difference from the photoemission onset to the low energy

cutoff. For a positive electron affinity the low energy cutoff will be determined by the vacuum level, and emission from the conduction band minimum will not be detected. A schematic of photoemission spectra of a semiconductor with a NEA or a positive electron affinity is shown in Fig. 2.[6] The electron affinity (χ) or the presence of a negative electron affinity can be deduced from the width of the spectrum (W) as follows:

$$\chi = h\nu - E_g - W \quad \text{for a positive electron affinity} \qquad (1)$$

$$0 = h\nu - E_g - W \quad \text{for a negative electron affinity}$$

where $h\nu$ is the photon energy and E_g is the bandgap. It is evident that for a positive electron affinity, the value of the electron affinity can be deduced from the measured width of the spectrum. It needs to be emphasized, however. that the absolute value of the electron affinity for a NEA surface cannot be measured by means of photoemission spectroscopy.

By carefully measuring the spectral width one can determine whether the low energy emission originates from the conduction band minimum. In fact recent measurements have indicated emission that extends several tenths of an eV below the conduction band minimum. Bandis and Pate have ascribed this emission to excitons for the C(111) surface exhibiting a NEA.[7] It was found that the observation depended on the band bending near the surface. For flat band and upward band bending, exciton emission was observed while for downward band bending the emission was ascribed to electrons in the band.[7] The band bending may be due to states in the bandgap that cause Fermi level pinning. Another possibility is H passivation of the boron acceptors near the surface which will lead to different band bending for the different regions on the surface.[8]

Photoemission spectroscopy can also be used to determine the position of the surface Fermi level. For a grounded sample, the Fermi level of the sample will be the same

as that of the metal sample holder. And the Fermi level of the metal can easily be determined. This measurement can be employed to monitor the semiconductor Fermi level or to detect band bending. We note that care must be taken to avoid photovoltage shifts that may be observed in highly resistive samples.

Secondary electron emission (SEE) is another technique that can also be used to characterize the surface.[9,10,11] To facilitate the secondary emission experiments, the sample is exposed to a monochromatic source of high energy electrons. Typically accelerating voltages are 1-5kV. Electron-hole pairs are generated in the conduction and valence band of the semiconductor by the incident electrons. The electrons then move toward the surface and may be emitted as described in the photoemission process. In general, SEE is less surface sensitive than UPS, since the electrons generated in SEE are distributed deeper in the sample than those from UPS. A typical application of this technique is to measure the electrons emitted per incident electron. We note that the gain is obtained since a single high energy electron can excite numerous electron-hole pairs that can be emitted and detected. A negative electron affinity surface will enhance the emission of electrons.

An alternative measurement is to obtain the energy spectrum of the emitted electrons. Similar to photoemission spectra, a negative electron affinity would be indicated by the presence of a sharp low energy peak in the SEE spectra (corresponding to the one displayed in Fig. 2). The energy of the incident electron beam depends on the work function of the electron gun relative to that of the target material. This effect makes an analysis based on a measurement of the width of the spectrum more difficult than for UPS.

It also needs to be emphasized that since the electron-hole pairs are excited deep in the sample in comparison to UPS any band bending at the semiconductor surface is going to influence the SEE gain. In particular, an upward band bending will inhibit electron transport to the surface while a downward band bending will sweep electrons towards the

surface. Therefore UPS is more suitable and easier to use than SEE to determine the electron affinity of a semiconductor surface.

15.3.3 Surface Termination Effects due to Molecular Adsorbates on Diamond

Different surface terminations can shift the position of the bands with respect to the vacuum level and, therefore, induce a NEA or remove it. Such changes have been found for molecular surface adsorbates. Different molecular surface adsorbates result in changes of the surface dipole layer, and for a wide bandgap material the surface dipole layer can lead to a positive or negative electron affinity. For example hydrogen has been reported to induce a NEA on the diamond (111) surface.[1,2,6,12] More recently a NEA effect has been shown for the hydrogenated C(100) and C(110) surfaces.[8,13,14,15] In comparison oxygen leads to a dipole such that a positive electron affinity is observed on these surfaces.

Various surface treatments designed to remove non-diamond carbon result in an oxygen terminated surface. These include acid etching and an electrochemical etch process. It has been suggested that different treatments can lead to different bonding configurations (bridge - versus double bonding) of oxygen on the C(100) surface.[8,16,17] Vacuum annealing of cleaned C(100) to 500°C does not remove a significant portion of the oxygen from the surface as detected by means of Auger electron spectroscopy (AES) indicating that most of the oxygen was chemisorbed. The UPS spectra of as-loaded samples as well as those heated to 500°C showed a positive electron affinity. A value for the electron affinity of 1.5 eV was detected following the 500°C anneal. Following an anneal to about 1000°C the amount of oxygen on the surface dropped below the detection limit of the AES instrument. A reconstructed (2x1) LEED pattern appeared. In addition, the width of the

11

UPS spectra increased to ~ 15.7 eV, and a sharp low energy feature appeared indicating a NEA following the last annealing step. It was presumed that the surface was hydrogen terminated after this step.

Annealing the (100) surface to greater than 1150°C results in the desorption of the remaining hydrogen, and a clean surface is obtained that exhibits a 2x1 reconstruction. The UPS measurements of this surface display a positive electron affinity. Fig. 3 shows UV photoemission spectra for oxygen terminated, clean and hydrogen terminated diamond (100) surfaces. The NEA character of the H-terminated surface is evident in both the width of the spectrum and the presence of the sharp peak at the lowest binding energy.

Further hydrogen plasma exposed surfaces exhibited a NEA and 2x1 reconstruction indicating a monohydride termination. These studies were actually preceded by ab initio calculations suggesting a NEA for the monohydride terminated 2x1 reconstructed (100) surface [13]. The same theoretical studies reported a positive electron affinity for the clean 2x1 surface in agreement with experimental observations. In fact the difference in the electron affinity between the two surfaces was ~3eV indicating the magnitude of the change in the surface dipole.

Diamond (110) surfaces also exhibited oxygen termination after surface cleaning, and they were found to have a positive electron affinity as evidenced by means of UPS.[8] Subsequent to annealing the samples to 700°C, the oxygen concentration on the surface dropped to below the detection limit of the AES instrument. The low energy cut off of the UPS spectrum shifted to lower energies, indicating a reduction of the electron affinity. In addition, a sharp low energy peak attributed to a NEA appeared. A 800°C anneal removed the sharp NEA feature, and the width of the spectrum was reduced by 0.7 eV. Exposing the surfaces to a H plasma resulted in the re-appearance of the NEA characteristics. Employing a 800°C anneal the NEA could be removed again. The results for the (100) and (110) hydrogen terminated, oxygen terminated and clean surface are summarized in Table

12

1.

In Fig. 4 a schematic of the band alignments for a clean and hydrogen terminated diamond surface is shown, with band bending consistent with p-type doping. Hydrogenation changes the surface dipole layer. This causes a shift of the bands with respect to the vacuum level. Overall, oxygen is bonded the strongest to diamond (100) - and the weakest to diamond (111) surfaces, and the bond strength of O on the C(110) surface falls between these two values.[8] Following this approach, according to the annealing temperatures necessary to remove a NEA from a hydrogen terminated diamond surface, the hydrogenation of the diamond (100) surface appears to be the most stable while that of the (110) the least thermally stable, while C(111) falls in-between.[8]

Recent work has explored the effects and stability of deuterium bonding on diamond. Deuterium termination has been found to induce a NEA effect comparable to hydrogen, but the surface was stable to a higher temperature than a hydrogen terminated surface. Future studies should focus on the stability of the hydrogen or deuterium terminated surfaces and the interactions with oxygen.

15.3.4 Metallic and Oxide Layers on Diamond

Deposition thin films of metal is another possibility of inducing an effective NEA on diamond. It has been known for many years that low workfunction metals like Cs can induce NEA type characteristics on III-V semiconductors like GaAs. High efficiency photocathodes are based on such structures. Cs deposition on diamond has also been demonstrated to induce a NEA effect.[18,19] Since diamond has a large bandgap other, higher workfunction metals may be suitable to establish a NEA. A few years ago Ti [20] or Ni [21] have been found to induce a NEA on diamond (111). More recently NEA characteristics have been reported for Co, Cu or Zr on diamond (100), (111) and (110)

surfaces.[16,22-24] Fig. 5 shows a summary photoemission spectra of diamond surfaces coated with a few monolayers of these metals. A sharp peak indicative of a NEA is detected for all these spectra.

Deposition of a thin metal layer was found to change the effective electron affinity of the diamond surface. This effect is illustrated in Fig. 6 for Ti on the clean diamond surface. This structure was found to exhibit a NEA. The effective electron affinity for a thin metal layer on the diamond can be modeled in terms of two interfaces: the vacuum-metal and the metal-diamond. Equation (2) gives an expression for the effective electron affinity.

$$\chi_{eff} = (\Phi_M + \Phi_B) - E_G \qquad (2)$$

In this model a lower Schottky barrier height would result in a lower effective electron affinity, and this is consistent with the experimental results. For each metal the surface termination prior to metal deposition appears to be have a significant effect on the Schottky barrier height. For metals deposited on clean surfaces lower values for the Schottky barrier height and a greater likelihood of inducing a NEA are expected than for metals on non adsorbate free surfaces. Fig. 7 shows the band diagrams for Ti on a clean or a H-terminated diamond surface. Metal - diamond interfaces exhibiting a NEA have a lower Schottky barrier height than those exhibiting a positive electron affinity.

In Fig. 8 the experimentally (by means of UV photoemission) measured Schottky barrier heights are plotted vs. the metal workfunction.[24] It is indicated whether a NEA or a positive electron affinity was observed by means of UV photoemission. The experimantal data are also compared to equation (2).

Of the results obtained to date the most significant may be the observation of a NEA of Co and Zr on diamond (100), (111) and (110).[23,24] These films have been shown to

be uniform with little tendency to the islanding that has been observed for Ni and Cu.[23,25] It has also been demonstrated that the NEA was stable following air exposure for Cu, Co and Zr on diamond surfaces.[23,24] It has also been demonstrated that Ti-oxide induces a NEA on diamond (111) surfaces.[26]

15.4 Field Emission

15.4.1 Description of Field Emission Process

Most practical applications of electron emitters will require field induced emission. The actual field emission process from a semiconductor combines four effects (1) electron supply to the semiconductor, (2) transport through the semiconductor, (3) emission into the vacuum and (4) transport in vacuum to the anode. For an ideal structure with a negative electron affinity and a low resistance contact and semiconductor, the electron emission would be limited by space charge in the vacuum, the I-V characteristics would exhibit a $V^{3/2}$ dependence. For the case of a positive electron affinity and low resistance contact and semiconductor, the material would essentially respond as a metal and Fowler-Nordheim characteristics would be obtained. For an intrinsic semiconductor with a negative electron affinity the emitted current would be limited by electron injection into the semiconductor, and since this may be a tunneling process, Fowler-Nordheim characteristics may also be obtained. In this situation the current could also be limited by space charge effects in the semiconductor.

Field (electron) emission is not only a promising means to develop intense/controlled electron currents for a variety of devices. It is also a powerful and rapid research tool to study the mechanisms of electron emission, whatever the means of stimulation.

15

15.4.2 Techniques

The field emission characteristics of diamond surfaces are most commonly determined by using either a movable probe as an anode that can be stepped toward the surface of the specimen [23,27,28], a large flat anode kept at a certain distance (typically in the mm range) away from the sample by a spacer [29,30,31] or a probe that is used to scan across the surface.[31,32]   In the case of a movable anode the current voltage characteristics are measured at different distances. The distances are of the order of a few μm to a few 10 μm. To avoid emission from edges of the anode it often has a spherical shape with a diameter of a few mm. This method has the advantage that the emission can be determined as a function of distance which may be important for the characterization of rough or highly resistive samples. The technique employing the large anode has the advantage of exploring the entire surface area. Also if a fluorescing anode is used, the location of various emission sites can be observed. However in this method field emitted electrons with KeV energies are striking the anode and can desorb material from there. In turn the cathode is exposed to KeV energy ions from the anode. As a result the surface of the cathode gets contaminated and discharges can occur. These effects can lead to changes in the emission behavior. Also when employing a phosphor screen as an anode problems can occur since small amounts of phosphor powder may get pulled to the cathode due to the high electric fields. Using a concentrated binder with the phosphor may alleviate this problem. To avoid the problems with phosphor an indium tin oxide (ITO) layer can be used instead.[30] But a higher voltage is necessary for ITO, increasing the likelihood of changing the emission characteristics due to contamination of the cathode surface or the striking of a discharge. Also the spatial resolution of an ITO coated anode is worse due to the higher energy of the electrons hitting the anode. Overall, to avoid discharging effects, a

16

better vacuum needs to be maintained for the technique using a large anode kept at a large distance, than for a method employing a probe in close proximity of the emitting surface. A variation of this technique is the arrangement of an anode grid close to the surface facilitating the field emission. A collector screen can then be placed at a larger distance away from the sample surface.[31] The third technique employs a tip that is placed a few μm from the emitting surface and can be scanned across this surface. This method can be useful for correlating the distribution of emission sites with the surface morphology.

15.4.3 Field Emission Results from Diamond

There have now been several studies that indicated electron emission from flat diamond surfaces at relatively low fields. Kordesh and co-workers used field emission microscopy and showed a uniform emission from p-type diamond films at low fields (~3V/μm).[33] Using a scanning probe with μm resolution Talin et al. reported turn on fields of 3V/μm for about 1/2 of the surface area of nanocrystalline diamond.[32] Raman spectra of the emitting areas exhibited lower quality sp^3 bonding. Latham and co-workers also measured field emission from flat diamond films at fields of ~20V/μm. They also showed that the emission was from very specific point sites that may be correlated with defect structures rather than sharp features of the films.[34] Zhu and co-workers found a correlation with specific sp^2 bonding structures in the films.[28,35] As shown in Fig. 9 higher electric fields were necessary for emission from high quality p-type diamond than for defective or particulate diamond.[35] To date, most of these measurements were on p-type boron doped diamond. Miskovsky and Cutler have analyzed the emission possibilities from these films and suggested that states in the bandgap must be present and participate in the emission process to account for the low field emission from the p-type material.[35,36]

It has been found that thin metal films of Cu, Co, or Zr on natural crystal p-type

17

diamond resulted in a decrease in the field emission threshold compared to the oxygen terminated diamond surface. The results of the field emission threshold and electron affinity showed a similar trend in which the field emission threshold decreased as the electron affinity decreased.[36,37]

Prins has developed a method of ion implanting diamond to produce nominally n-type material.[38,39] Geis and co-workers have used this technique to fabricate a diamond cold cathode emitter structure based on an all diamond p-n junction.[40] Carbon ion implantation was employed to induce the n-type like characteristics in diamond. Fig. 10 shows a schematic of the cold cathode device.[40] The diode current vs. applied voltage for the carbon ion implanted diode was compared to that of an aluminum Schottky contact on p-type diamond. Fig. 11 illustrates the measured I-V characteristics for both structures.[40] Geis et al. have also obtained field emission from nitrogen doped diamond.[27] There are several possible configurations for nitrogen impurities in diamond, but for nitrogen occupying single substitutional sites, the impurities exhibit n-type characteristics with a donor level at ~1.7eV below the conduction band. These materials also showed low field emission, but the spatial variation has not been reported.

There have also been several recent studies of the field emission from diamond coated field emitters (i.e. pointed structures) made of silicon or metals.[33,41-46] In these measurements the electron emission was found at significantly lower fields than from uncoated surfaces. However many studies have shown nonuniform growth of CVD diamond on the sharp tips. In particular, distinct diamond particles have been observed on the tips as shown in Fig. 12.[47] Possible explanations for the electron emission based on negative electron affinity diamond surfaces or based on different radii of silicon tips underneath the diamond have been proposed.[46,47] The growth of nanocrystalline CVD diamond has been reported.[32,48-50] This material may be suitable for coating field emitter tips (or flat emission surfaces) with a smooth, uniform diamond layer. Preliminary

results of emitter tips coated with nanocrystalline diamond indicate a significant (3-fold) reduction in field emission threshold compared to the uncoated tips.[51]

It is evident that the complicated processes involved in field emission have impeded the advancement of our understanding of these measurements. The measurements themselves require care, and in fact, some early reports may have been dominated by artifacts attributable to poor vacuum or other effects due to the high fields present in the measurements. The residual gas in the measuring chamber results in a background ion current. Moreover the strong fields can sometimes cause a plasma to ignite and severely damage the surface. Following arc discharges, crater formations and molten areas with debris were observed in the surface of CVD diamond and amorphous carbon films.[52] A corresponding improvement in the field emission has been reported. It has been suggested that this improvement may be due to the formation of protrusions or sharp edges that could act as emission sites.

Even under ultrahigh vacuum conditions sputtering or desorption of material from the anode can be caused due to the high energy electrons emitted from the sample. This may result in deposition of material on the sample or in a destructive discharge. Positive ions will be accelerated towards the surface of the sample and may damage the surface, even if no plasma discharge occurs. One way to avoid these damaging effects is the use of lower voltages. This implies that smaller distances need to be employed to facilitate the field emission. At smaller distances, effects due to the surface roughness of the sample or the anode become more significant.

Simultaneous field emission and photoemission measurements from a (111) 1x1:H p-type natural diamond surface were reported by Bandis and Pate.[53] From these experiments it was found that the field emitted electrons originated from the valance band maximum. Fig. 13 shows the electron energy distribution for the simultaneous field emission and photoemission measurements.[53] In contrast the photoemission process

involved emission from the conduction band. For boron doped polycrystalline diamond films Glesener and Morrish found no dependence of the field emission on temperature, and they also suggest that the electrons originate from the valence band.[54] To date there have been no studies that confirmed field electron emission originating from the conduction band of diamond. Most studies on emission from diamond may be explained as electron emission from the valence band or from defect states. Only UV photoemission spectroscopy measurements, where UV radiation is employed to excite electrons from the valence band into the conduction band, have studied the electron emission from the conduction band.

Supplying electrons to the conduction band remains a significant problem. Ideally a shallow n-type dopant could solve this issue. Then a field emission structure could consist of a highly n-type doped region at the electron injecting contact and lower doping in the bulk of the material. Nitrogen is a substitutional dopant and has been reported in concentrations of 10^{19} cm^{-3} in diamond.[55] It has been reported that high concentration of incorporated nitrogen could enhance electron emission.[27,29] According to theoretical [56] and experimental [57,58] studies the "shallowest" level of nitrogen is located about 1.5-2.1 eV below the conduction band minimum. Extremely high nitrogen concentrations would be necessary to get a n-type doping effect and to facilitate the electron injection at the back contact. There has been great difficulty in incorporating such large amounts of nitrogen into diamond. Alternatively, roughening of the interface at the contact may help circumvent this difficulty and lower the effective barrier for electron injection due to field enhancement at the rough interface.[27] A novel approach for CVD diamond deposition has been reported to enable the incorporation of large amounts of nitrogen into diamond.

Phosphorus is another substitutional impurity that may be a potential n-type dopant in diamond. For substitutional phosphorus an activation energy of 0.20 eV has been calculated.[56] Also the value for the equilibrium solubility of phosphorus in diamond is

expected to be low.[56]   A doping effect due to a shallow level of phosphorus has also been found experimentally.[59]   N-type conductivity and a corresponding value of 0.20-0.21 eV for the activation energy was measured for phosphorus implanted into high purity type IIa natural diamond.[60]   Deeper levels with activation energies of 0.84-1.16 eV have been calculated [61,62] and measured experimentally.[63,64]

Lithium is an interstitial impurity in diamond and may be another potential n-type dopant.   A donor level of 0.1 eV below the conduction band minimum has been reported.[56]   But diffusion of lithium would result in an undesirable deterioration of the doping characteristics over time.[56]   Also the solubility of lithium in diamond is predicted to be low.

The commonly observed non-uniformity in the emission from diamond surfaces [27] may also be a significant obstacle for the potential use of diamond in emission devices. This implies that the emission site density becomes a crucial characteristic to determine whether an emitting surface is suitable for practical applications.   It has been estimated that an emission site density of the order of at least 10^6 to 10^7 sites/cm^2 is necessary for applications as field emission displays.[65]   Figures 14 a) and b) show emission images of a carbon layer for 6.5 V/μm and 10 V/μm.[65]   An image of a differently prepared carbon film for 10V/μm is shown in Figure 14 c).[65]   A comparison of the emission site density for these two carbon films can be seen in Figure 15.[65]   A reported increase in emission site density of CVD diamond films has been attributed to a pre-growth treatment.[66] Figure 16 shows the emission site density of a treated and an untreated area.[66].

However, uniform emission from diamond films has been reported in one study using a novel surface electron microscope that was operated in the field emission mode.[67,68] The instrumental limit for the field was 2V/μm.   Since typical threshold fields in other reports are about an order of magnitude higher it may be conceivable that there is a uniform emission at low fields which becomes non-uniform at higher fields.   This issue will need to

21

be resolved for most applications.

15.5 Conclusions

It is now evident that obtaining a true NEA is possible for diamond. Results indicate that a positive electron affinity is obtained for both adsorbate free and oxygen terminated surfaces. However, a NEA is obtained for hydrogen termination of all low index surfaces. It is interesting to note that more than a decade transpired between the first NEA measurements on (111) surfaces, and the discovery of a NEA for H-terminated (100) surfaces. It may be that the more tenacious bonding of oxygen to the (100) surface was at least partially responsible for the delayed observation.

The properties of thin metal layers on diamond have also indicated that these may be suitable for obtaining a NEA surface. The model used to describe this effect is based on two interfaces - the vacuum-metal interface and the metal-diamond interface. Within this model, to lower the effective electron affinity of this structure it is necessary that the metal-diamond interface changes the surface dipole of the diamond. A particularly encouraging result was found for Zr deposited on clean diamond surfaces and surfaces with oxygen or H adsorbates. Here it appears that the Zr displaces the oxygen or hydrogen termination.

The field emission is the most complicated and potentially least understood measurement. The combined photo and field emission measurements indicate that field emission from p-type diamond originates from electrons in the valence band. Similarly, many studies have indicated lower field emission thresholds for diamond with significant defect density. This again suggests that the emission does not involve electrons in the conduction band. While conduction band emission has been suggested for nitrogen doped diamond, the role of defects needs to be further explored.

The two most pressing questions in the field emission studies is verifying if field

22

emission from the conduction band has been obtained and understanding the non-uniform emission. Certainly the development of a process to obtain shallow n-type doping would go a long way to solving each of these problems.

Even without the development of an n-type dopant it seems likely that diamond with defects may substantially improve the emission character of various pointed and flat surfaces. Preparation of actual device structures may be necessary to determine if the current status is sufficient for the applications.

Diamond has been the wide gap semiconductor that has been most often considered for emission applications, but recent studies have indicated a negative electron affinity for both BN and AlN. It appears that n-type doping of these materials is also problematic, but other approaches may be available for supplying electrons to the conduction band.

This review has neglected some aspects of field emission. The most notable of these may be the beneficial effects of roughness and interfaces. However, the complexities of the processes may provide a real challenge for future research.

References

1. Himpsel, F.J., J.A. Knapp, J.A. van Vechten, D.E. Eastman, Phys.Rev. B20 (1979), 624.

2. Pate, B.B. (1986). Surf. Sci. **165**, 83.

3. R. Gomer, *Field Emission and Field Ionization*, (Harvard Press, Cambridge, 1961).

4. Zangwill, A (1988), *Physics at Surface*, (Cambridge).

5. R. J. Nemanich, P. K. Baumann, M. C. Benjamin, S. W. King, J. van der Weide, R. F. Davis, "Negative electron affinity surfaces of aluminum nitride and diamond" Diamond Related Mat. 5 (1996) 790-796

6. J. van der Weide, R. J. Nemanich, "Argon and Hydrogen Plasma Interactions on Diamond (111) Surfaces; Electronic States and Structure", Appl. Phys. Lett. 62 (1993), 1878-1880.

7. C. Bandis and B.B. Pate, "Electron Emission Due to Exciton Breakup from Negative Electron Affinity Diamond," Phys. Rev. Lett. 74 (1995), 777.

8. P. K. Baumann, R. J. Nemanich, "Surface Cleaning, Electronic States and Electron Affinity of Diamond (100), (111) and (110) Surfaces" submitted to J. Appl. Phys.

9. G.T. Mearini, I.L. Krainsky, J.A. Dayton, Jr., Y. Wang, C.A. Zorman, J.C. Angus, D.F. Anderson, "Stable secondary electron emission from chemical vapor deposited diamond films coated with alkali-halides," Appl. Phys. Lett. 66 (1995), 242-244.

10. D.P. Malta, J.B. Posthill, T.P. Humphreys, R.E. Thomas, G.G. Fountain, R.A. Rudder, G.C. Hudson, M.J. Mantini, and R.J. Markunas, "Secondary electron emission enhancement and defect contrast from diamond following exposure to atomic hydrogen," Appl. Phys. Lett. 64 (1994), 1929-1931.

11. R.E. Thomas, T.P. Humphreys, C. Pettenkofer, D.P. Malta, J.B. Posthill, M.J. Mantini, R.A. Rudder, G.C. Hudson R.J. Markunas, "Influence of surface

terminating species on electron emission from diamond surfaces," Mat. Res. Soc. Symp. Proc. vol 416 (1996), 263-274.

12. R. J. Nemanich, L. Bergman, K. F. Turner, J. van der Weide and T. P. Humphreys, "Properties of Interfaces of Diamond," Trieste Semiconductor Symposium on Wide-Band-Gap Semiconductors, Physica B 185 (1993), 528-538.

13. van der Weide, J., Z. Zhang, P. K. Baumann, M. G. Wensell, J. Bernholc and R. J. Nemanich, "Negative Electron Affinity Effects on the Diamond (100) Surface", Physical Review B 50 (1994), 5803-5806.

14. van der Weide, J., and R. J. Nemanich,"Angle-Resolved Photoemission of Diamond (111) and (100) Surfaces; Negative Electron Affinity and Band Structure Measurements", J. Vac. Sci. Technol. B 12, 2475-2479 (1994).

15. P. K. Baumann and R. J. Nemanich, "Negative Electron Affinity Effects on H Plasma Exposed Diamond (100) Surfaces", Diamond Relat. Mat. 4 (1995), 802-805.

16. P.K. Baumann, T.P. Humphreys and R.J. Nemanich, "Comparison of Surface Cleaning Processes for Diamond (100)," Mat. Res. Soc Symp. Proc. vol 339 (1994), 69-74.

17. M.J. Rutter and J. Robertson, submitted to Phys. Rev. B

18. M.W. Geis, J.C. Twichell, J. Macaulay, and K. Okano, "Electron field emission from diamond and other carbon materials after H2, O2, and Cs treatment," Appl. Phys. Lett. 67 (1995), 1328-1330.

19. O. M. Küttel, O. Gröning, E. Schaller, L. Diederich, P. Gröning and L Schlapbach, "Electron field emission from a caesiated NEA diamond (100) surface: an activation concept" Diamond Related Mat. 5 (1996) 807-811.

20. J. van der Weide, and R. J. Nemanich, "Schottky Barrier Height and Negative Electron Affinity of Titanium on (111) Diamond," J. Vac. Sci. Technol. B 10 (1992), 1940-1943 .

21. J. van der Weide, and R. J. Nemanich,"Influence of Interfacial Hydrogen and Oxygen on the Schottky Barrier of Nickel on (111) and (100) Diamond Surfaces", Physical Review B49 (1994), 13629-13637.

22. P.K. Baumann, R.J. Nemanich, "Characterization of Co-diamond (100) interfaces: electron affinity and Schottky barrier," Appl. Surf. Sci.104/105 (1996) 267; "Negative electron affinity effects and Schottky barrier height measurements od metals on diamond (100) surfaces," Mat. Res. Soc. Symp. Proc. vol. 416 (1996), 157-162. and "Electron Affinity and Schottky Barrier Height of Metal - Diamond Interfaces": J. Vac. Sci. Tech. B 15(4) (1997) in press

23. P. K. Baumann, R. J. Nemanich, "Characterization of Copper - Diamond (100), (111) and (110) Interfaces: Electron Affinity and Schottky Barrier" submitted to Phys. Rev. B

24. P.K. Baumann and R.J. Nemanich, "Electron Affinity and Schottky Barrier Height of Metal Diamond (100), (111) and (110) Interfaces" submitted to J. Appl. Phys.

25. P. K. Baumann, T. P. Humphreys, R. J. Nemanich, K. Ishibashi, N. R. Parikh, L. M. Porter and R. F. Davis, "Epitaxial Cu Contacts on Semiconducting Diamond", Diamond Related Mat. 3 (1994), 883-886.

26. Bandis, C., D. Haggery, and B.B. Pate, "Electron emission properties of the NEA (111) 2x1 diamond -TiO interface," Mat. Res. Soc. Symp. Proc. vol. 339 (1994), 75-81.

27. M.W. Geis, J.C. Twichell, N.N. Efremow, K. Krohn, and T.M. Lyszczarz, "Comparison of electric field emission from nitrogen-doped, type Ib diamond, and boron doped diamond," Appl. Phys. Lett. 68 (1996), 2294-2296.

28. W. Zhu, G.P. Kochanski, S. Jin, L. Seibles, D.C. Jacobson, M. McCormack and A.E. White, Appl. Phys. Lett. 67 (1995), 1157-1159.

29. K. Okano, S. Koizumi, S.R.P. Silva, G.A.J. Amaratunga, "Low-threshold cold

cathodes made of nitrogen-doped chemical-vapour-deposited diamond," Nature 381 (1996), 140-141.

30. P.V. Latham, <u>High Voltage Vacuum Insulation</u>, edited by R.V.Latham, Academic Press, San Diego (1995).

31. Z. Feng, I. G. Brown and J. W. Ager III, "Electron emission from chemical vapor deposited diamond and amorphous carbon films observed with a simple field emission device" J. Mater. Res. vol. 10 (1995) 1585-1588.

32. A. A. Talin, L. S. Pan, K. F. McCarthy, T. E. Felter, H. J. Doerr and R. F. Bushah, "The relationship between the spatially resolved field emission characteristics and the raman spectra of a nanocrystalline diamond cold cathode" Appl. Phys. Lett. vol. 69 (1996) 3842-3844.

33. C. Wang, A. Garcia, D.C. Ingram, M. Lake, and M.E. Kordesch, Electron. Lett. 27 (1991), 1459.

34. N.S. Xu, R.V. Latham and Y. Tzeng, Electron. Lett. 29 (1993), 1596.

35. W. Zhu, G.P. Kochanski, S. Jin, "Electron field emission properties of diamond" Mat. Res. Soc. Symp. Proc. vol. 416 (1996) 443-448.

36. Z.-H. Huang, P.H. Cutler, N.M. Miskovsky and T.E. Sullivan, Appl. Phys. Lett. 65 (1994), 2562-2565.

37. N.M. Miskovsky, P.H. Cutler and Z.-H. Huang, J. Vac. Sci. Technol. B 14 (1996) 2037 - 2040.

38. J. Prins, "Ion-implantation and diamond - some recent results on growth and doping", Thin Solid Films vol. 212 (1992) 11-18.

39. J. Prins, "Ion-implanted structures and doped layers in diamond". Materials Science Reports vol. 7 (1992) 271-364.

40. M. Geis, N. Efremow, J. Woodhouse, M. Mcalese, M. Marchywka, D. Socker and J. Hochedez, "Diamond cold cathode", IEEE Elect. Dev. Lett. vol. 12 (1991) 456-

459.

41. N. S. Xu, Y. Tzeng and R. V. Latham, J. Phys. D: Appl Phys., 26 (1993) 1776

42. J. Liu, V. V. Zhirnov, A. F. Myers, G. J. Wojak, W. B. Choi, J. J. Hren, S. D. Wolter, M. T. McClure, B. R. Stoner and J. T. Glass, J. Vac. Sci Technol. B 13 (1995) 422.

43. W. B. Choi, J. J. Cuomo, V. V. Zhirnov, A. F. Myers and J.J. Hren, Appl. Phys. Lett. 68 (1996) 720.

44. W. B. Choi, J. Liu, M. T. McClure, A. F. Myers, V. V. Zhirnov, J. J. Cuomo and J.J. Hren, J. Vac. Sci. Technol. B 14 (1996) 2050.

45. V. Raiko, R. Spitzl, B. Aschermann, D. Theirich, J. Engemann, N. Puteter, T. Habermann, G. Müller, Thin Solid Films 290-291 (1996) 190.

46. E. I. Givargizov, V. V. Zhirnov, A. N. Stepanova, E. V. Rakova, A. N. Kieselev, P. S. Plekhanov, "Microstructucture and field emission of diamond particles on silicon tips" Appl. Surf. Sci. vol. 87/88 (1995) 24-30.

47. V. V. Zhirnov "On the cold emission mechanism of diamond coated tips" J. de Physique IV vol. 6 (1996) C5-107 - C5-112.

48. D. M. Gruen, X. Pan, A. R. Krauss, S. Liu, J. Luo, C. M. Foster "Deposition and characterization of nanocrystalline diamond films" J. Vac. Sci. Technol. A vol. 12 (1994) 1491-1495.

49. D. M. Gruen, C. D. Zuiker, A. R. Krauss, "Diamond films grown from fullerene precursors" Proc. of the SPIE - The International Society for Optical Engineering vol. 2530 (1995) 2-13.

50. C. D. Zuiker, A. R. Krauss, D. M. Gruen, X. Pan, J. C. Li, R. Csencsits, A. Erdemir, C. Bindal, G. Fenske, "Physical and tribological properties of diamond films grown in argon-carbon plasmas" Thin Solid Films vol. 270 (1995) 154-159.

51. A. R. Krauss, personal communication

52. O. Gröning, O. M. Küttel, E. Schaller, P. Gröning and L Schlapbach, "Vacuum arc discharges preceding high electron field emission from carbon films" Appl. Phys. Lett. vol. 69 (1996) 476-478.

53. C. Bandis, B.B. Pate, Appl. Phys. Lett. 69 (1996) 366.

54. J. W. Glesener and A. A. Morrish, Appl. Phys. Lett. 69 (1996) 785.

55. T. Evans, The properties of natural and synthetic diamond, edited by J.E. Field, Academic Press, London (1992) 259

56. S.A. Kajihara, A. Antonelli, J. Bernholc and R. Carr Phys. Rev. Lett. vol. 66 (1991) 2010.

57. W.J.P. Enkevort and E.H. van Versteegen, J. Phys.: Condens. Matter vol. 4 (1992) 2361.

58. R.G. Farrer, Solid State Commun., vol. 7 (1969) 685

59. K. Okano, H. Kiyota, T. Iwasaki, T. Kurosu, M. Iida and T. Nakamura, New Diamond Science and Technology, MRS Int. Conf. Proc. (1991) 917

60. J.F. Prins, Diamond Relat. Mat. vol. 4 (1995) 580

61. K. Jackson, M.R. Pederson and J.G. Harrison, Phys. Rev. B, vol. 41 (1990) 12641

62. V.V. Tokiy, N.D. Samsonenko, D.L. Savina and S.V. Gorban, Proc of the 2nd Int. Conf. on the Applications of Diamond Films and Related Materials, August 25-27, 1993, Japan, MYO, Tokyo p.757

63. M.I. Landstrass, M.A. Plano, D. Moyer, S.P. Smith, and R.G. Wilson, Diamond Materials, Electrochemical Society (1991) 574.

64. M. Kamo, H. Yarimoto, T. Ando and Y. Sato, New Diamond Science and Technology, MRS Int. Conf. Proc. (1991) 637.

65. A. A. Talin, B. F. Coll, E. P. Menu, J. Markhalm and J. E. Jaskie, "Carbon cathode requirements and emission characterization for low-voltage field emission displays", Proc. of the 1st Specialist Meeting on Amorphous Carbon, Cambridge, UK, July 31,

1997.

66. Z. L. Tolt, R. L. Fink and Z. Yaniv, "Electron emission from patterned diamond flat cathodes", Proc. of IVMC'97, J. Vac. Sci. Technol. A

67. J.D. Shovlin and M.E. Kordesch, "Electron emission from CVD diamond and dielectric breakdown," Appl. Phys. Lett. 65 (1994), 863-865.

68. J.D. Shovlin, M.E. Kordesch. D. Dunham, B.P. Tonner, W. Engel, "Synchrotron radiation photoelectron emission microscopy of CVD diamond electron emitters", J. Vac. Sci. Technol. A13 (1995), 1111-1115.

Table 1. The UPS spectral width for different diamond (100) and (110) surface terminations. The electron affinity or presence of a NEA is deduced using Eq. 1.

Surface	UPS Spectral Width (eV)	Electron Affinity (eV)
C(100):H	15.7	NEA
C(100) (clean)	15.05	0.65
C(100):O	14.2	1.50
C(110):H	15.7	NEA
C(110) (clean)	15.1	0.60
C(110):O	14.3	1.40

Physical Properties Affecting
the Workfunction or Electron Affinity

**Atomic Levels
(Referenced to
Vacuum)**

Ionization Energy

Charge Density

Surface Dipole

Positive Charge

Negative Charge

z →

**Modification of
Surface Dipole**

1) Reconstruction

**2) Adsorbed
Molecules**

3) Steps

**Adsorbed
Molecules**

Surface Dipole

- → +
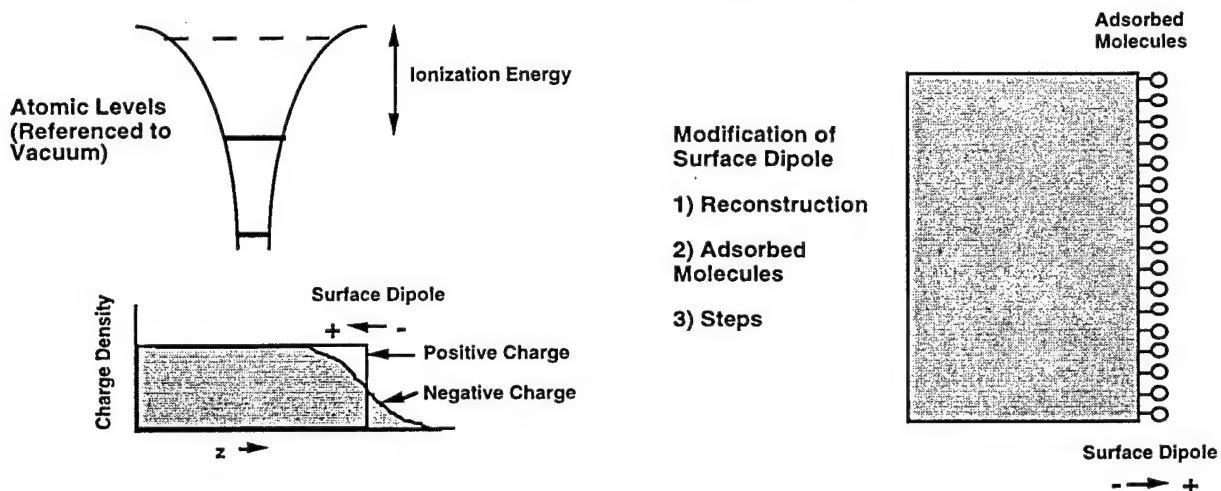
Fig. 1 A representation of the effects which contribute to the work function (or electron affinity) of any material. While the atomic levels are an intrinsic property of the material, changes in the surface bonding can substantially affect the work function or electron affinity.
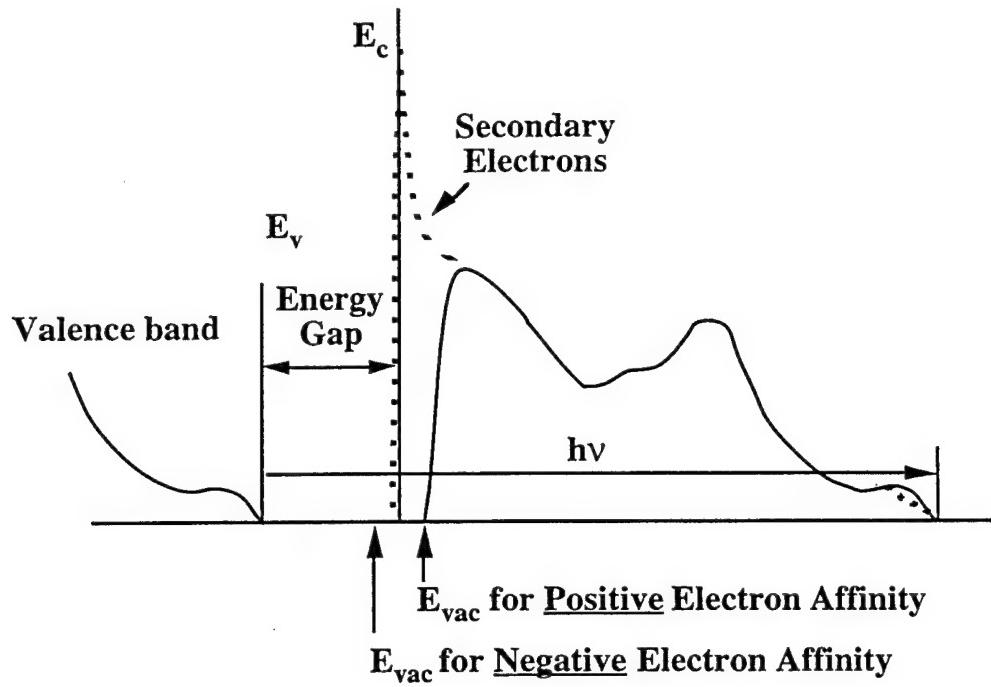
Fig. 2 A schematic of how NEA affects the photoemission spectra. For a NEA surface the spectra is broadened to lower kinetic energy and a peak due to quasi thermalized electrons is detected also at the lowest kinetic energy. [6]
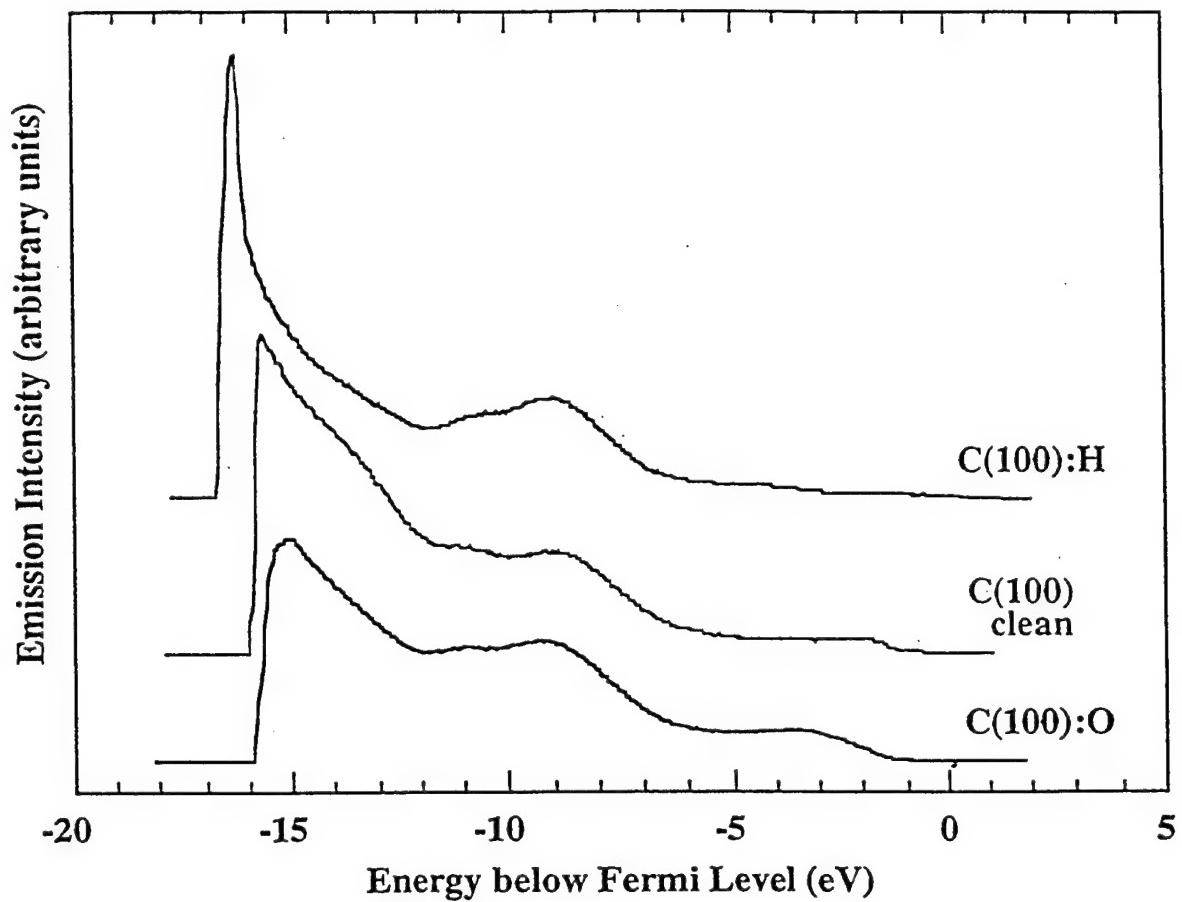
Fig. 3 Photoemission spectra of oxygen, hydroterminated and clean diamond (100) surfaces. The broadening of the spectral width and the sharp feature at high (negative) binding energy (i.e. low kinetic energy) are indicative of a NEA.

Electronic Bands Near the Surface

Fig. 4 The band alignments of clean diamond and H terminated diamond surfaces. Note that the figures have been aligned at the vacuum level.
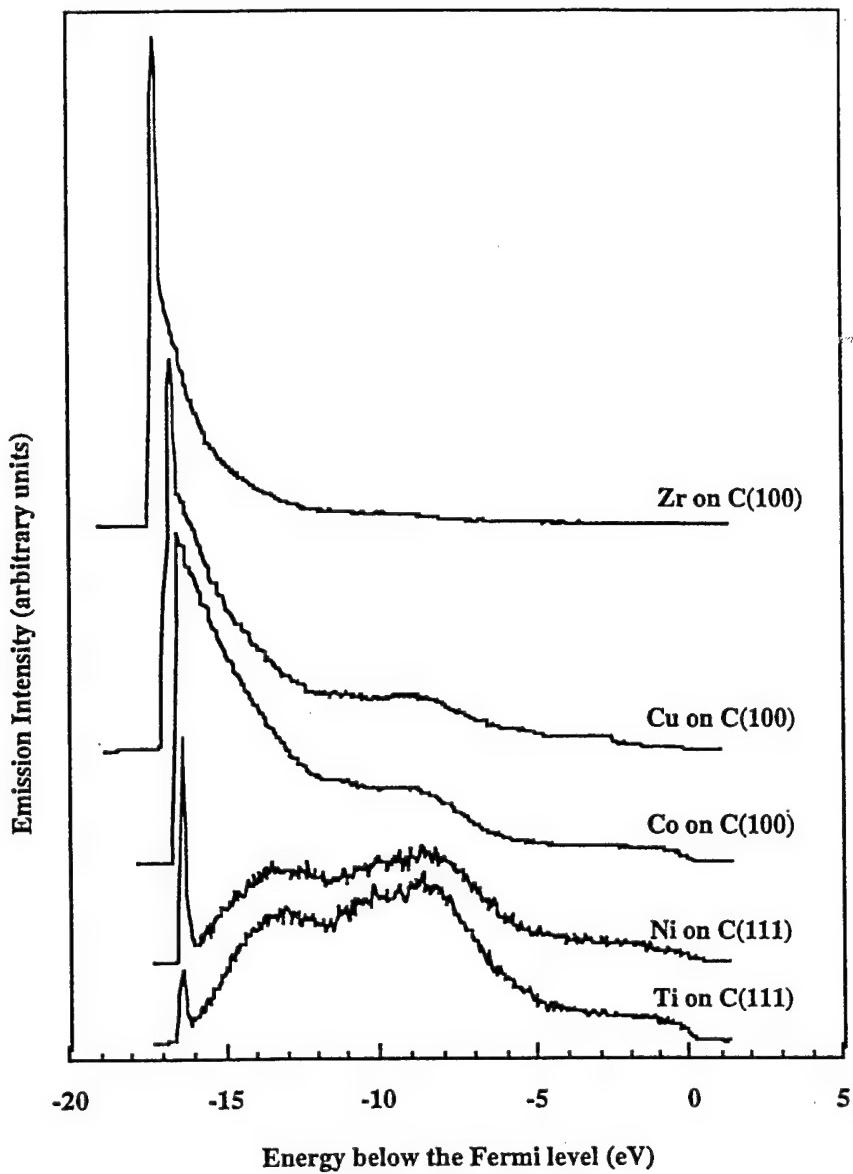
Fig. 5. UV-photoemission spectra of diamond surfaces with thin metal overlayers. The metal thicknesses correspond to several monolayers.

Electronic Bands Near the Surface
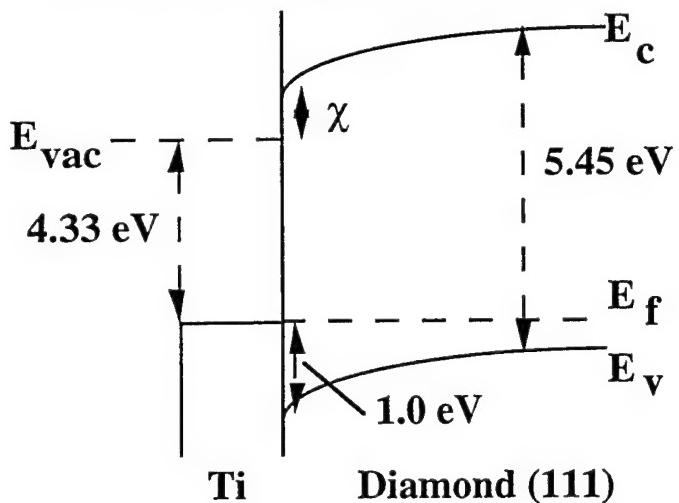
The electron affinity is then given by

$$\chi = (\psi_m + \varphi_p) - E_g$$

$$= (4.33 + 1.0) - 5.45$$

$$= -0.1 \text{ eV}$$

Fig. 6. The band structure at the surface of diamond with a thin metal coverage. The electron affinity can be deduced from the Schottky barrier, the metal work function and the diamond bandgap. The numbers for Ti on diamond (111) are illustrated.

Electronic Bands Near the Surface

Clean Diamond (111)

H / Diamond (111)
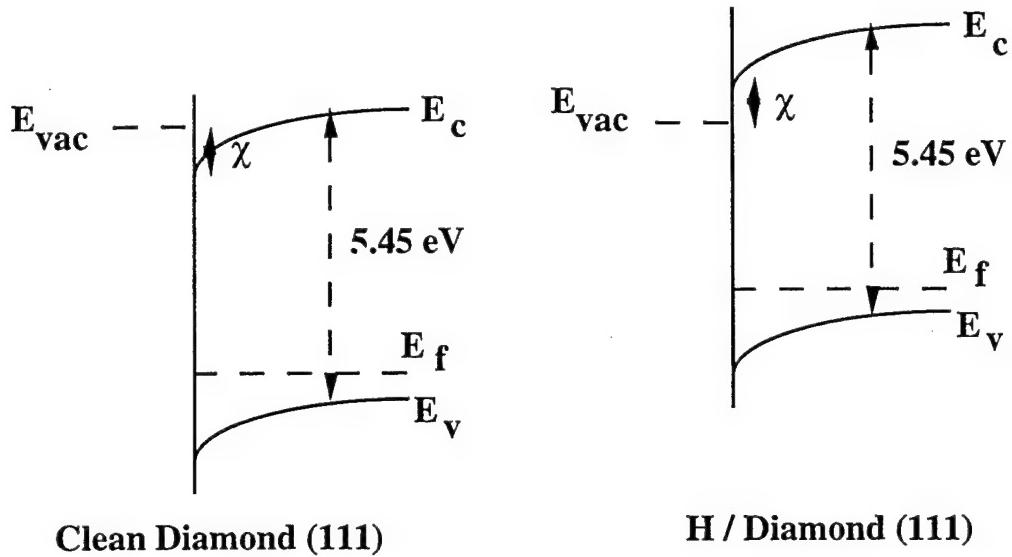
Fig. 7 Schematic band diagram of Ti on clean and H-terminated diamond surfaces. Note that the figures have benn aligned at the vacuum level.

Fig. 8. Diagram of the Schottky barrier height vs. metal workfunction for Ti, Zr, Cu, Co and Ni. The dashed line represents the limit for which a NEA is expected for metal - diamond interfaces according to equation (2). Thus a NEA is expected for data points below this dashed line and a positive electron affinity for those above. The experimental data for Ti, Zr, Cu, Co and Ni are plotted. The filled markers correspond to an experimentally observed NEA and the empty markers indicate an experimentally observed positive electron affinity.[24]
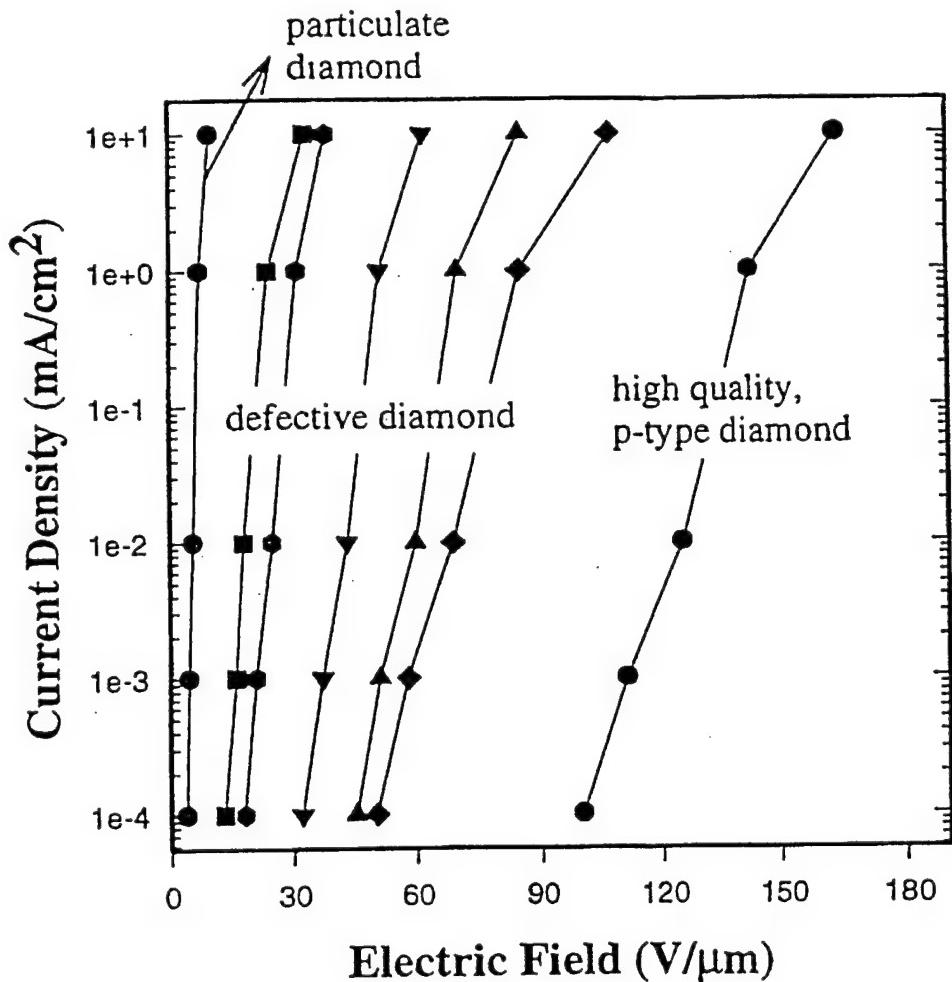
Fig. 9 Plot of the emission current densities vs. field required for emission from p-type

CVD diamond, defective undoped diamond and nanometer size diamond powder.[33]
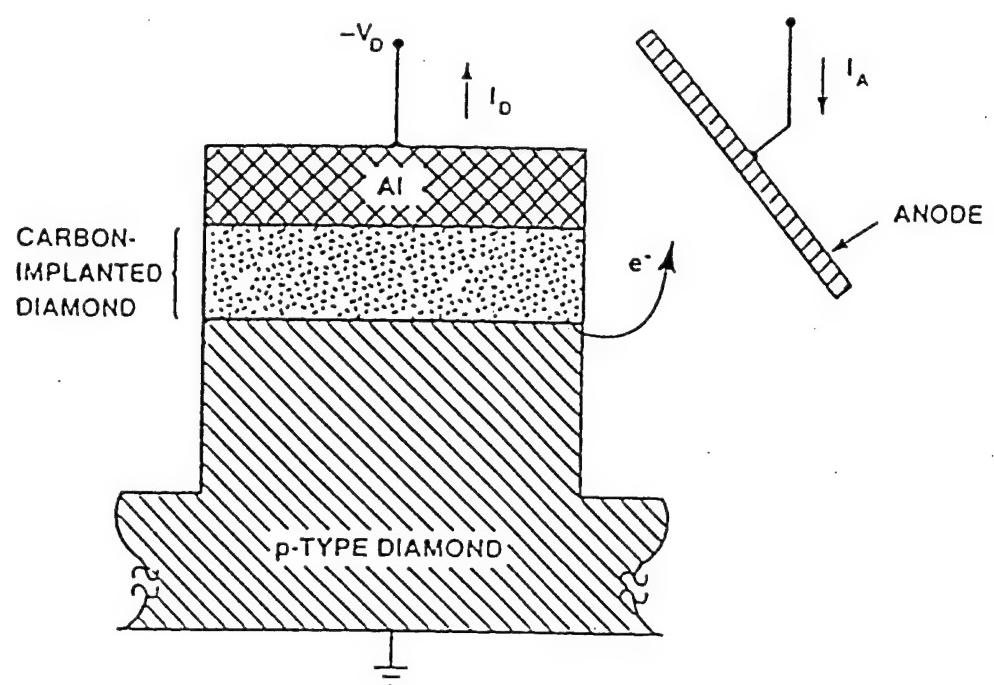
Fig. 10 Schematic of the cold cathode based on an all diamond p-n junction.[38]

41
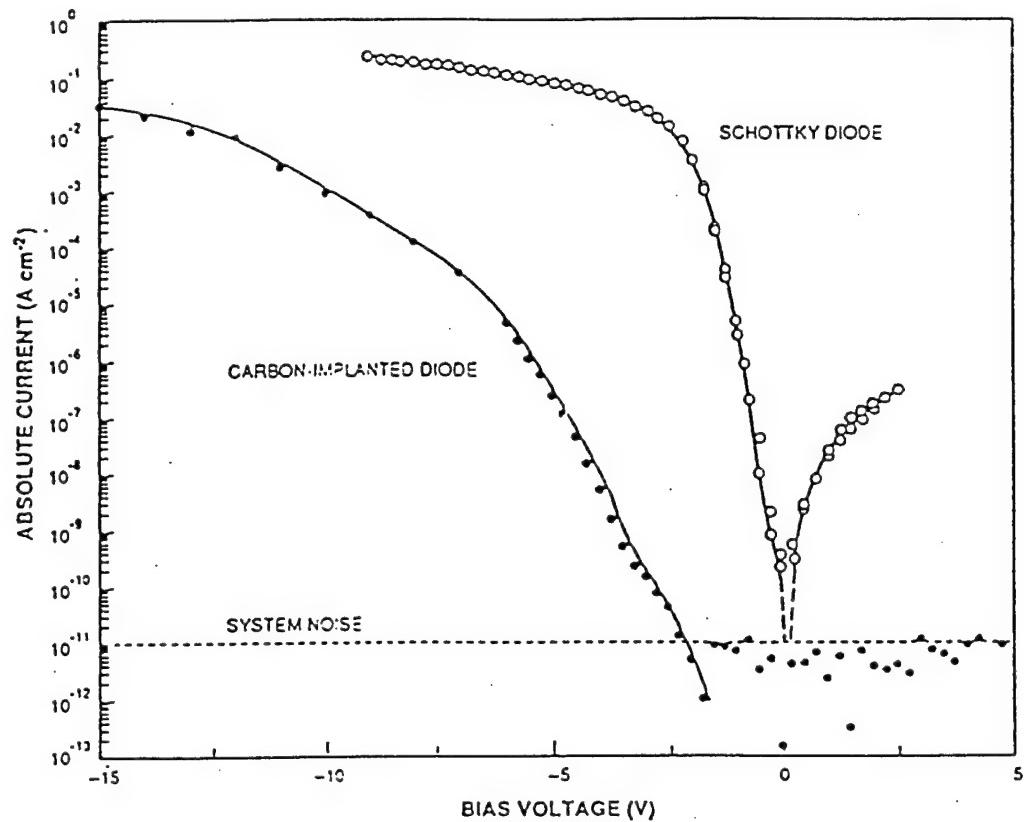
Fig. 11 Diode current vs. applied voltage for the carbon ion implanted diode was compared to that of an aluminum Schottky contact on p-type diamond. Illustrates the measured I-V characteristics for both structures.[38]
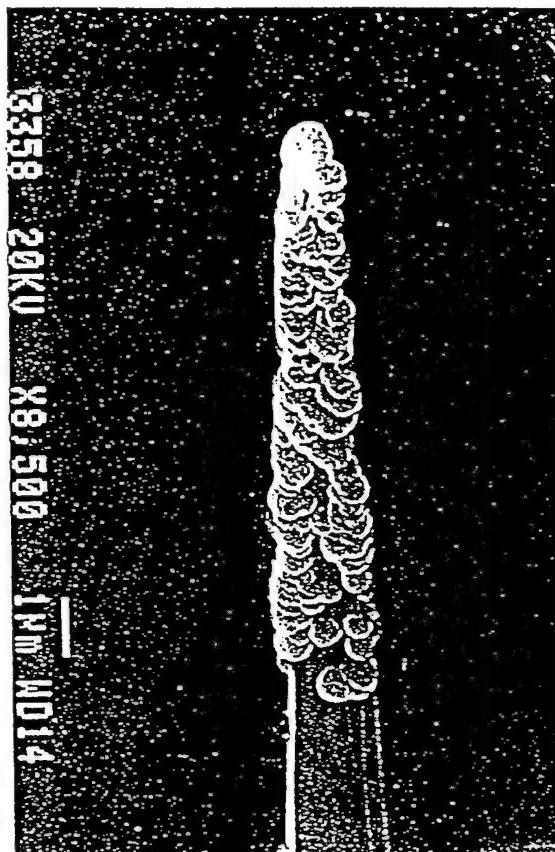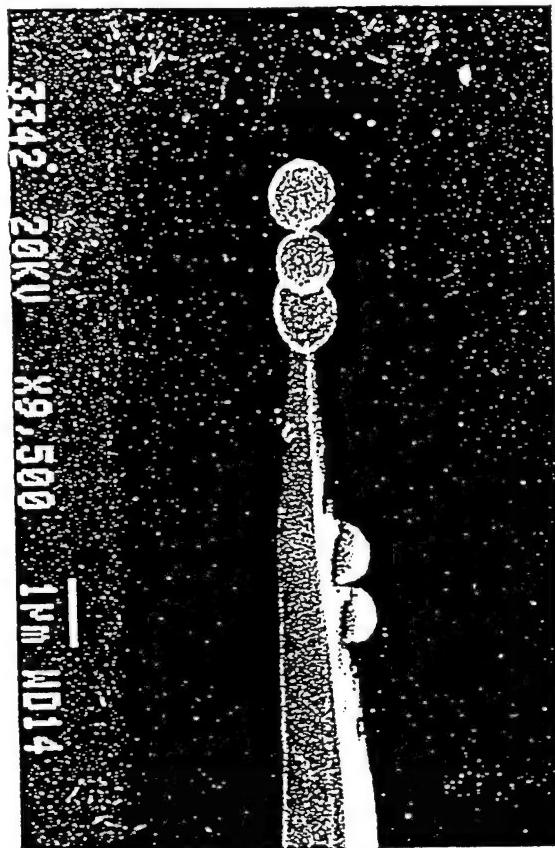
Fig. 12. Distinct diamond particles on the silicon tips. a) single-particle-type diamond coverage. b) Conglomerate of diamond particles.[45]
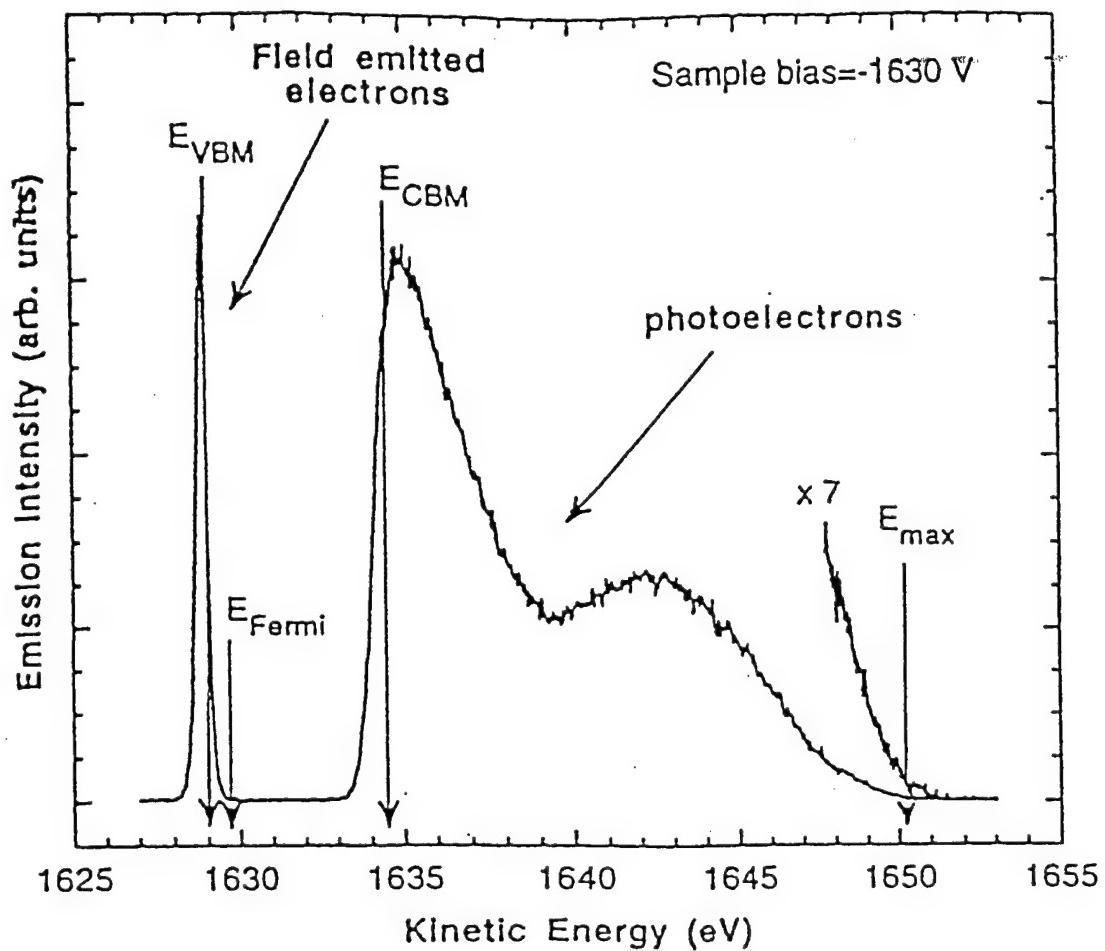
Fig. 13 Electron energy distribution for the simultaneous field emission and photoemission measurements on C(111)1x1:H. The energy position of the field emitted electrons corresponds to the valence band maximum.[51]
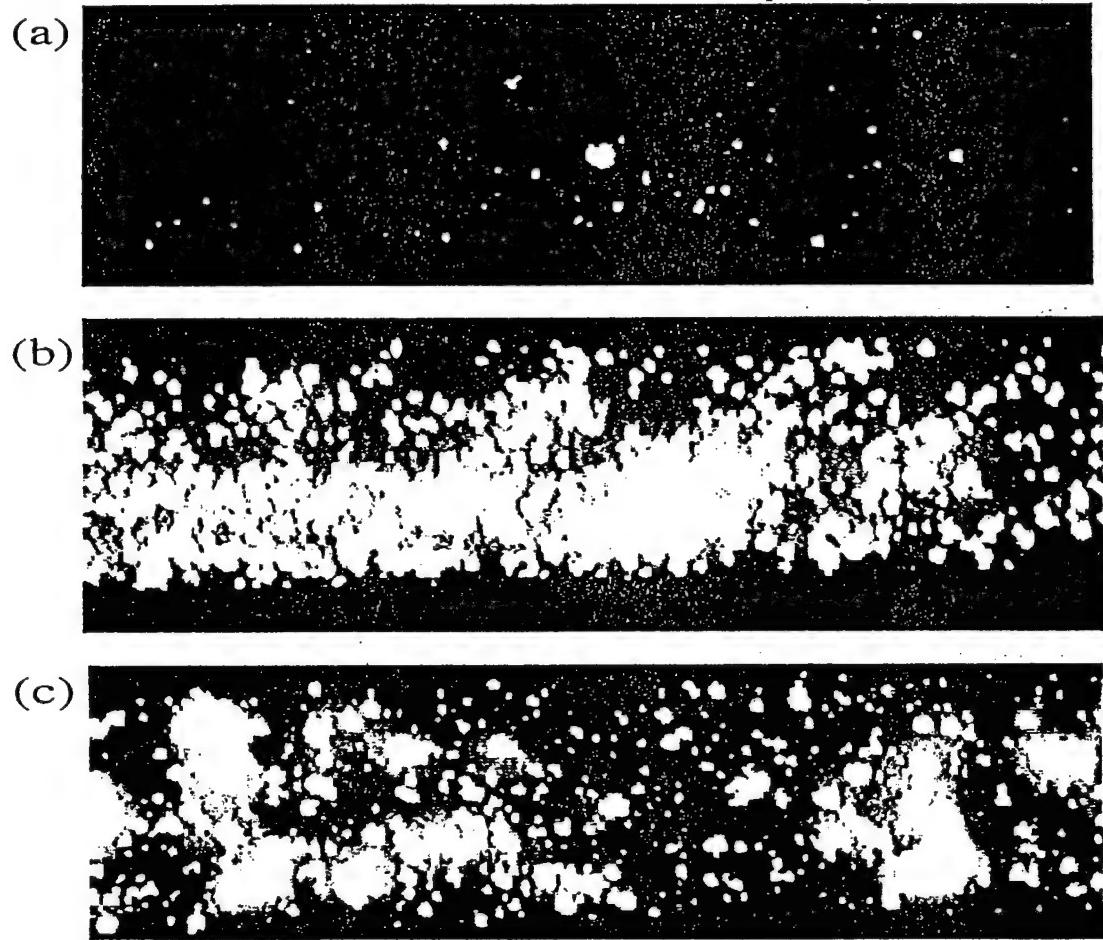
Fig. 14 Field electron emission from a) cathode 1, 6.5 V/μm, b) cathode 1, 10 V/μm, c) cathode 2, 10V/μm.[65]
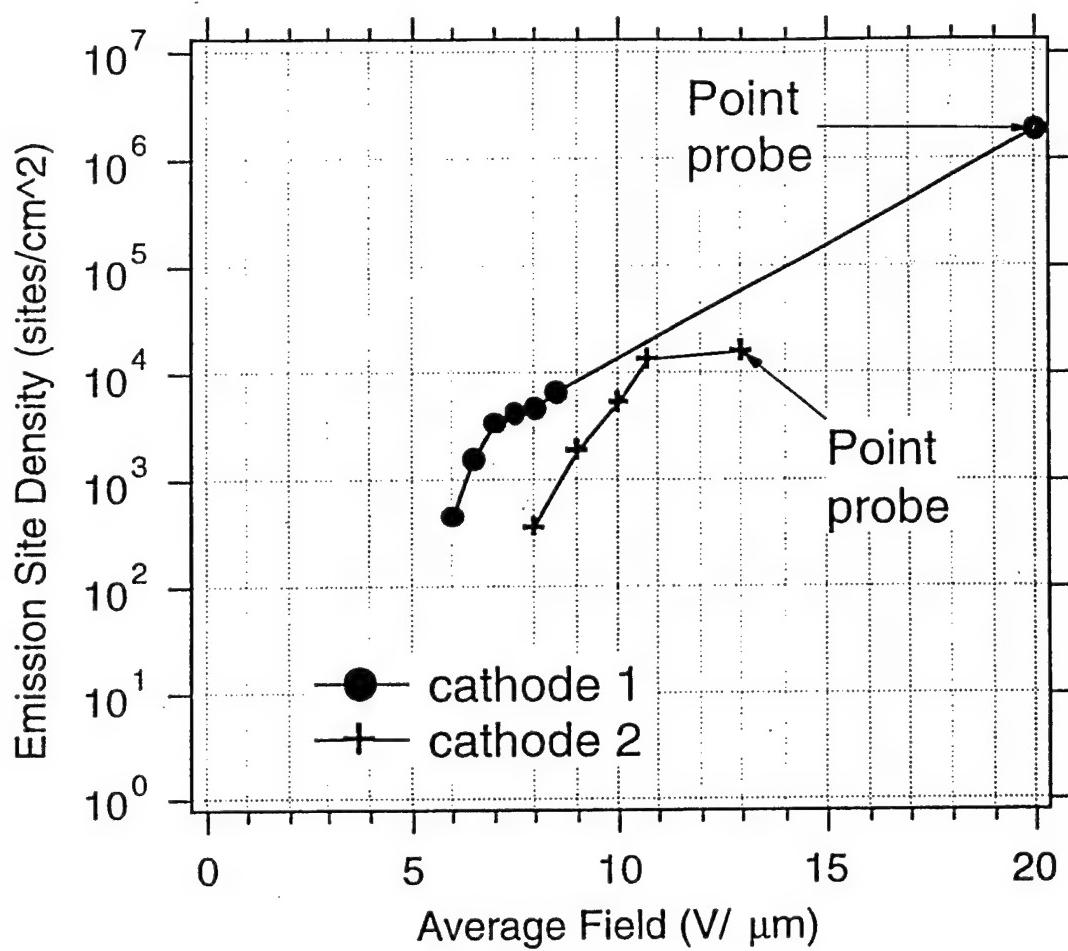
Fig. 15 Plot of the emission site density versus average field for cathodes 1 and 2.[65]
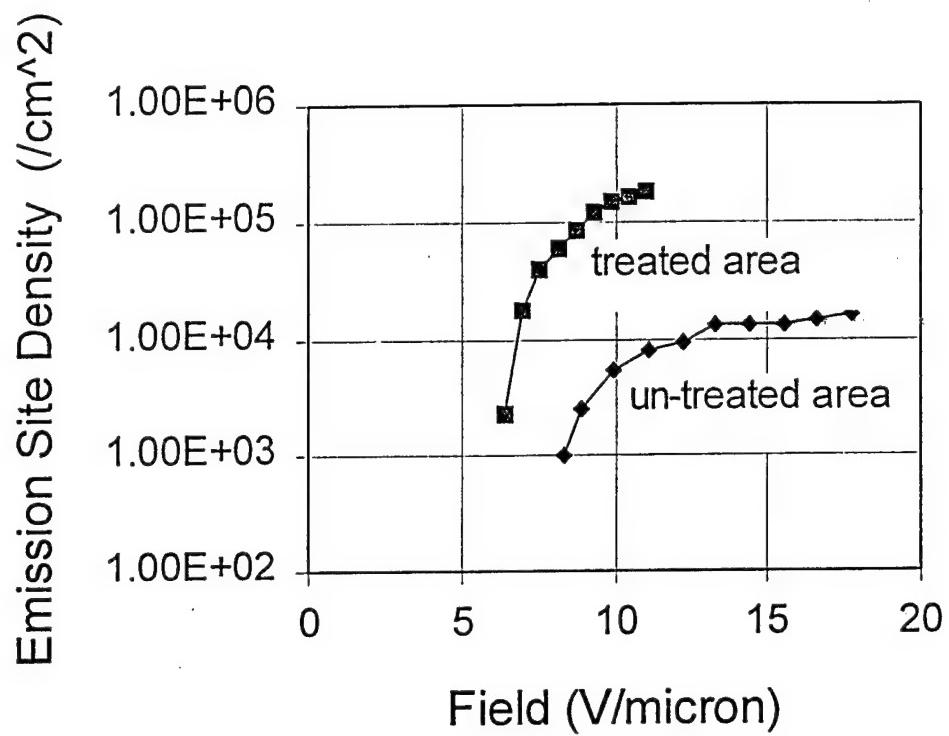
Fig. 16 Plot of the emission site density versus field for treated and untreated areas of the substrate.[66]

III. Electron Emission Mechanism from Cubic Boron Nitride Coated Molybdenum Emitters

B.L. McCarson, R. Schlesser, M.T. McClure, and Z. Sitar
Department of Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695-7919

Abstract

The energy distribution of field emitted electrons from intrinsic cubic boron nitride (c-BN) coated Mo tips was analyzed in an effort to determine the origin of the emitted electrons. Voltage dependent field emission energy distribution (V-FEED) spectra were collected from the Mo emitters under UHV conditions both before and after being coated. Emission current at a given voltage increased an order of magnitude with the c-BN coated emitters relative to bare emitters. The energy of field emitted electrons from the c-BN coated emitters was linearly dependent upon the applied voltage. Extrapolation of V-FEED data from c-BN coated emitters to flat-band condition evidenced that the electrons were emitted from the conduction band minimum of the c-BN coating at the c-BN/vacuum interface.

Introduction

In recent years, a significant amount of research effort has been applied toward enhancing the performance of microstructured field emitters for use in the field of vacuum microelectronics and displays.[1,2] Research on the use of molybdenum as a field emitter for applications, such as flat panel display devices, has demonstrated that Mo is a promising material for this task.[3] However, to meet some application requirements, emitters must be coated with a chemically and mechanically stable material that will either not change or improve the emission characteristics of the device. This has stimulated an interest in coating emitters with wide bandgap materials. The use of wide bandgap materials as coatings for Mo tips has demonstrated an enhancement of the field emission current relative to bare Mo tips.[4,5] One such chemically inert and mechanically robust coating material is cubic boron nitride (c-BN). Recent UV-photoemission measurements have demonstrated that c-BN features negative electron affinity.[6] In addition, field emission studies on flat n-type c-BN have shown that considerable emission currents are attainable. These results make the prospect of c-BN as a coating material for field emitters promising.[7] Characterization of c-BN as a coating material for Mo emitters is the subject of this study. The following describes the methods employed and results of our field emission studies by voltage dependent field emission energy distribution (V-FEED) and simultaneous current versus voltage (I-V) measurements.

Experimental Procedure

Each Mo emitter was fabricated by electrochemically etching a Mo wire of 125 μm initial diameter. During the etching, the Mo wire was biased at +10 V DC and submerged in a concentrated potassium hydroxide solution with a platinum counter electrode. The tip morphology and radii of curvature were determined by Scanning Electron Microscopy (SEM). Typical radii of curvature were less than 100 nm. The c-BN coatings were applied by an electrophoretic procedure where the Mo tips were again biased at +10 V DC (also with a platinum counter electrode) and submerged in an ultrasonically prepared suspension of c-BN particles in ethanol. Typical c-BN particle size was approximately 100 nm. Coating thickness and uniformity were monitored by SEM. Coatings were usually a few hundred nanometers thick.

For analysis, the coated and bare emitters were secured in a holder so that each tip was centered 500 μm beneath a circular gate with an opening of 500 μm. The holder was then placed in an ultra high vacuum analysis chamber with a base pressure of 10^{-9} Torr and was aligned with the entrance lens of a hemispherical electron analyzer (VG Instruments, CLAM II). The electron analyzer recorded the energy distribution of the field emitted electrons which passed through the opening of the gate. Both the gate and analyzer were at ground potential while the emitter was biased negatively using a high voltage source with an integrated

picoammeter (Keithley 237). The energy of field emitted electrons was referenced to the Fermi level (E_F) of Mo using the following relationship (Fig. 1):

$$(E - E_F) = E_{kin} - eV + \Phi_A, \tag{1}$$

where V represents the potential applied between the gate and the tip, e the elementary charge, and E_{kin} the measured kinetic energy of field emitted electrons. The analyzer work function, $\Phi_A = 4.3$ eV, was calibrated using a method described elsewhere.[5] All measurements were performed at room temperature.

FEED spectra and I-V data from the bare Mo tips were collected prior to the coating procedure and after an *in vacuo* thermal desorption at 500°C for at least one hour. FEED spectra and I-V data were subsequently collected from the same emitters after completion of the electrophoretic coating procedure and after an additional *in vacuo* thermal desorption similar to that applied to the bare tips. The desorption step was needed to produce stable emission in both cases. Thermal annealing of diamond coated Mo tips has also been reported to stabilize emission.[5] However, this was in part due to the formation of a conductive carbide layer between the diamond and the Mo. No such conductive layer could be formed between c-BN and Mo through annealing. The more stable emission after annealing of bare and c-BN coated Mo tips arose probably due to the removal of Mo oxide that formed during the processing of the tips.

Results and Discussion

Figure 2 shows a typical FEED spectrum for a Mo field emitter displayed as a function of energy $(E - E_F)$. The count rate was proportional to the field emission current j. The solid
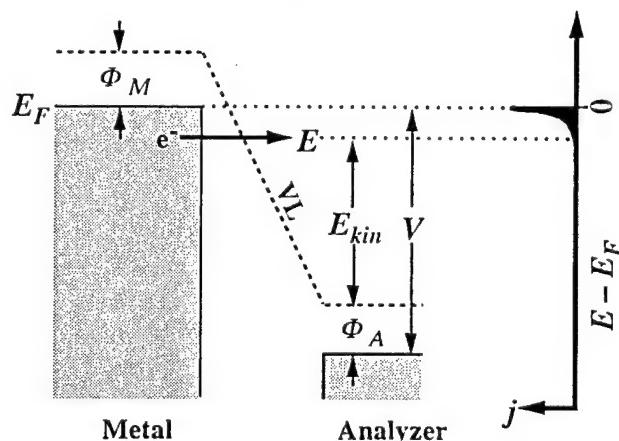
Figure 1.   Emission mechanism and theoretical FEED spectrum for a metallic field emitter. Electrons are emitted from below the metal's Fermi level E_F when a negative voltage (V) is applied between the emitter and the analyzer. Φ_M: metal work function; Φ_A: analyzer work function; j : field emission current; E_{kin}: measured kinetic energy of emitted electrons; VL: vacuum level.

50

curve represents a fit of the experimental data to a theoretical FEED spectrum, $j(E - E_F)$, convolved with a Gaussian instrument broadening function, $i(E - E_F)$. For metallic emitters the theoretical energy distribution of $j(E - E_F)$ is given by the following:[8]

$$ j(E - E_F) = \frac{e\pi(2m)^{1/2}}{h^3} \frac{f(E - E_F)}{(\Phi - (E - E_F))^{1/2}} \exp\left(-\frac{4}{3}\left(\frac{2m}{h^2}\right)^{1/2} \frac{(\Phi - (E - E_F))^{3/2}}{eF} \right), \qquad (2) $$

where e is the elementary charge, h the Planck's constant, m the electron mass, Φ the work function of the metal ($\Phi_M = 4.3$ eV)[9], $f(E - E_F)$ the Fermi-Dirac distribution, and F the electric field at the emission site. The local field, F, is connected to the applied voltage, V, via the field enhancement factor β, $F = \beta V$. The full width at half maximum of the instrument function was determined to be (0.15±0.02) eV at an analyzer pass energy of 2 eV. Fit parameters included the field enhancement factor, β, which was consistently on the order of 10^5 cm^{-1}, and a factor linking the observed count rates to the calculated field emission current j. The dashed line in Fig. 2 is the theoretical energy distribution calculated without consideration of instrument broadening. The energy $(E - E_F)$ of the electrons emitted from Mo was well defined and did not depend upon the potential applied between the tip and gate. This was expected since the applied field does not penetrate the metal. The only observable change in the emission characteristics of the bare Mo tips at varying voltages was an increase in the count rate at higher voltages.
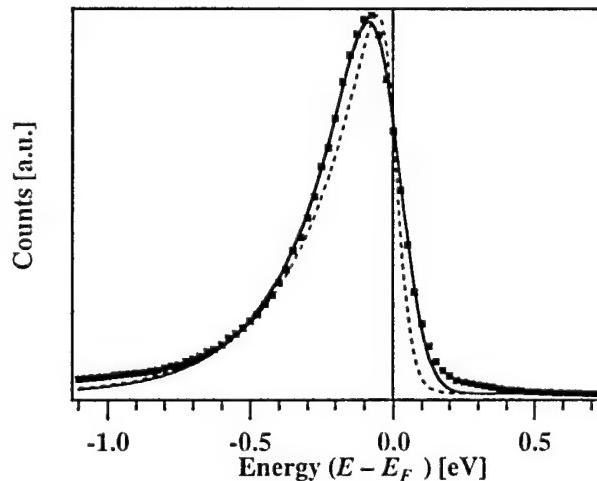

Figure 2.   FEED spectrum of a bare Mo emitter. Experimentally recorded data are represented by dots and were fitted to Eq. 2 and convolved with the instrument broadening function (solid line). The dashed line represents Eq. 2 without consideration of instrumentation broadening.

Subsequently, after a coating of c-BN was applied, the field emission current and energy distribution of electrons emitted from Mo tips were also measured. At a given voltage, the measured emission current increased by as much as one order of magnitude after the c-BN coating was applied. The energy $(E - E_F)$ of the electrons emitted from the c-BN coated tips decreased monotonically with the applied voltage as shown in Fig. 3.

The data in Fig. 4 describe the change in energy peak position $(E - E_F)$ with applied voltage of a c-BN coated Mo emitter. As can be seen, the peak position decreased linearly with the applied voltage. Extrapolation of the peak position data to flat-band condition (potential of
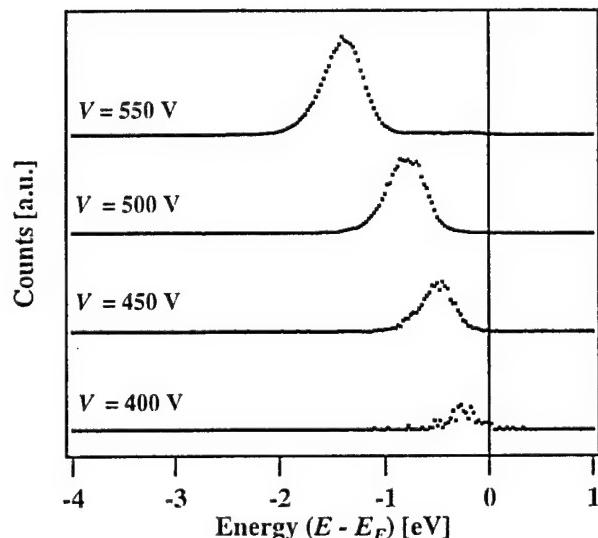

Figure 3.   FEED spectra of c-BN coated Mo emitter recorded at different applied voltages.
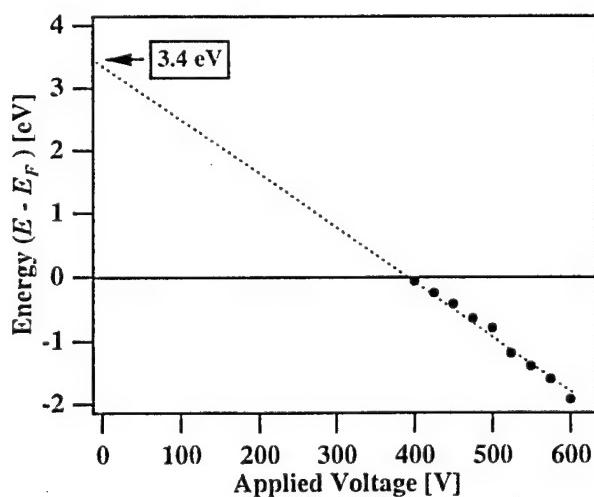
Figure 4.   FEED peak position variation as a function of the applied voltage. Linear extrapolation of data to a potential of 0 V yielded an intercept of 3.4 eV.

52

0 V) yielded an intercept of 3.4 eV. This indicated that the field emitted electrons originated 3.4 eV above the Fermi level of c-BN. As the c-BN powder used was nominally intrinsic, its Fermi level could be expected to be located close to midgap. C-BN has a bandgap of 6.2 eV,[7] thus the observed emission most probably originated from the CBM.

The approximately linear decrease in energy (E - E_F) with applied voltage can be explained by a similar model used to describe the behavior of diamond coated Mo emitters.[5] In this model, the change in energy is explained by considering the voltage drop across the dielectric coating (Fig. 5). Electrons are injected from the Fermi level (E_F) of Mo into the conduction band of the c-BN where they settle to the conduction band minimum (CBM). The observed emission thus originated from the CBM at the c-BN/vacuum interface. The original vacuum level (VL) is lowered by a voltage drop (ΔV) which is proportional to the applied voltage and the thickness of the layer, and inversely proportional to the dielectric constant of c-BN.

Conclusions

The emission mechanisms from bare Mo emitters and Mo emitters coated with intrinsic c-BN were analyzed by the V-FEED technique. The energy of field emitted electrons from c-BN was found to be linearly dependent upon the applied voltage, whereas the energy of electrons field emitted from bare Mo was not voltage dependent. Extrapolation of V-FEED data collected from c-BN coated emitters to flat-band condition revealed that the electron emission originated from the conduction band minimum. As such, the emission from intrinsic c-BN occurred by electron injection into the conduction band, complete thermalization of electrons to the CBM, and finally emission at the c-BN/vacuum interface.
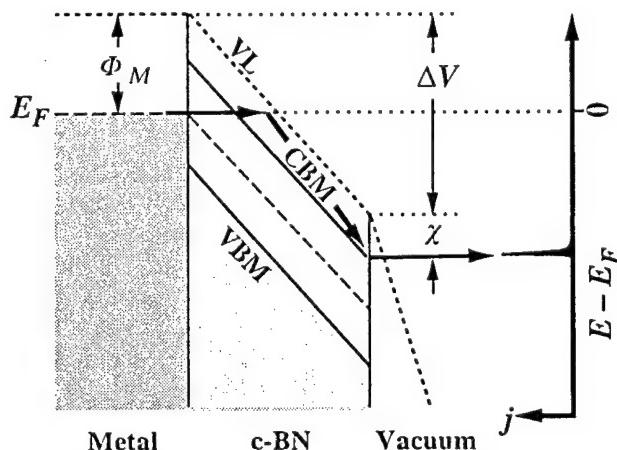
Figure 5. Schematic of emission mechanism for Mo coated with c-BN which consists of: electron injection into the conduction band, thermalization of electrons, and subsequent field emission from the conduction band minimum (CBM). Φ_M: Mo work function; χ: electron affinity of the wide bandgap material (assumed positive). Vacuum level (VL), CBM, and valence band maximum (VBM) are bent in the extracting electric field, so that a voltage drop ΔV occurs across the semiconductor layer.

53

References

1   I. Brodie, and C. A. Spindt, Adv. Electron. Electron Phys. **83**, 1 (1992).
2   P. R. Schwoebel, and I. Brodie, J. Vac. Sci. Technol. B **13**, 1391 (1995).
3   C. A. Spindt, J. Appl. Phys. **39**, 3504 (1968).
4   W. B. Choi, J. Liu, M. T. McClure, A. F. Myers, V. V. Zhirnov, J. J. Cuomo and J. J. Hren, J. Vac. Sci. Technol. B **14**, 2050 (1996).
5   R. Schlesser, M. T. McClure, B. L. McCarson, and Z. Sitar, Appl. Phys Lett. **70**, 1596 (1997).
6   M. J. Powers, M. C. Benjamin, L. M. Porter, R. J. Nemanich, R. F. Davis, J. J. Cuomo, G. L. Doll, and S. J. Harris, Appl. Phys. Lett. **67**, 3912 (1995).
7   R. W. Pryor, Mat. Res. Soc. Symp. Proc. **416**, 425 (1996).
8   P. Soven, E.W. Plummer, and N. Kar, Crit. Rev. in Solid State Sci. **6**, 111 (1976).
9   G. A. Haas, in *American Institute of Physics Handbook*, edited by D. W. Gray (McGraw-Hill, New York, 1972).

Growth and Characterization of AlN, GaN and InN and Alloys of These Compounds

IV. A Consistent Methodology for Calculating Surface Energies

Krzysztof Rapcewicz, Bin Chen, Boris Yakobson and J. Bernholc

Department of Physics, North Carolina State University, Raleigh, NC 27695-8202

A consistent approach to the calculation of the surface energy valid for all crystal systems is presented. Voronoi polyhedra are introduced and used in conjunction with the energy-density formalism of Chetty and Martin [Phys. Rev. B 45, 6074 (1992); ibid., p. 6089] to provide a methodology for the determination of surface energies. The surface energies of the unrelaxed, unreconstructed GaAs (001) and (111) surfaces are calculated as a test. As an example of the application of the formalism to a low symmetry system, the energies of selected (0001) surfaces of the wide-gap semiconductors GaN and SiC are determined.

INTRODUCTION

There has been tremendous interest and progress in the technology of crystal growth over the past few decades.[1] This has made relevant numerous theoretical questions related to crystal growth and the equilibrium behavior of surfaces. Indeed, intense efforts have been made to understand the energetics of such surfaces.[2] However, in spite of the substantial progress towards an understanding of the physics of surfaces, the surface energies must be calculated with respect to a reference surface, the reference being different for different surfaces except for specific high-symmetry cases. Since the surface energy plays an important role in the determination of the mode of growth, namely layer, island or layer-plus-island growth, the lack of a procedure for the calculation of the surface energy is unfortunate. Indeed, if meaningful *ab initio* predictions about the equilibrium crystal shape and preferred growth directions are to be made, it will be necessary to have a database of surface energies of reconstructed surfaces for different directions. This requires that the evaluations of the surface energies be consistent between calculations.

The situation can be summarized as follows: On the basis of general considerations, it can be shown that for crystals of sufficiently high symmetry, the surface energy can be calculated unambiguously. For the remaining systems, it can only be calculated consistently, *i.e.*, the surface energy is defined up to a gauge; this gauge is of no consequence in the determination, for instance, of equilibrium crystal shapes. Practically, however, the surface energy, when it is unambiguously defined, can be calculated absolutely with a slab calculation only when the two surfaces are crystallographically identical; most often, only a relative surface energy can be determined.

Three cases therefore exist: 1) the surface energy is unambiguously defined and can be determined using a slab calculation, the (001) surface of a zinc-blende semiconductor is such a case; 2) the surface energy is unambiguously defined, but slab calculations give only relative surface energies, the (111) surface of a zinc-blende semiconductor serves as a case in point; 3) the surface energy can only be defined consistently (the gauge must be set, once and for all, for the calculations), the (0001) surface of a wurtzite semiconductor is such an example.

Chetty and Martin[3-5] developed an approach that could be used to calculate the surface energies in the second case above. They introduced the concepts of the energy density and symmetry-adapted unit cell to provide a procedure for the calculation of the surface energy and applied their approach to the GaAs (111) surface and interfaces. However, symmetry-adapted unit-cells can be used only if the space group has sufficiently high symmetry, namely a center of symmetry, two axes of rotation, or an axis of rotation and a mirror plane not through this axis. For the third case, no methodology exists to calculate the surface energy.

We have generalized the energy-density approach of Chetty and Martin[3,4] to a consistent methodology for the calculation of surface energies for all systems *including those of case three above*. In this approach, the slab and bulk are considered to be built from blocks whose shape is determined by the symmetry of the bulk crystal. The "energy cost" of each block is evaluated and the "cost" of creating a surface, the surface energy, is obtained by summing up the cost of each block in the slab and subtracting the costs of the equivalent blocks in the bulk. The shape of the blocks is determined according to a geometric rule; thus the approach provides a consistent way of determining the surface energy that can be used for low symmetry systems. Further, the symmetric nature of the blocks ensures that symmetry-equivalent surfaces will have the same surface energy and hence the methodology will reproduce the results of slab calculations, and those of Chetty and Martin. The blocks can be chosen to be either neutral or charged, the former being our preferred choice because they are more intuitive.

The paper is organized as follows: a brief résumé of the surface energy begins the presentation, after which the standard calculations of the surface energy are discussed. In addition, a conceptual approach using only total energy calculations and scaling arguments, which is applicable to systems of sufficiently high symmetry, is presented. This approach clarifies the issues involved in unambiguously defining the surface energy, but it is not feasible at present. The energy-density approach

of Chetty and Martin is then reviewed together with the implementation of the energy-density formalism using symmetry-adapted unit cells. A generalization of the energy-density formalism using Voronoi polyhedra (VP) is introduced and applied to GaAs (001), GaAs (111), SiC (0001) and GaN (0001). The latter two cases provide examples of low symmetry systems in which symmetry-adapted unit cells cannot be used. The paper concludes with a summary.

I. THE SURFACE ENERGY

In this section we review the theory of the surface energy and the role that symmetry plays in the definition of the surface energy. In particular, for systems that are of low symmetry the surface energy cannot be defined absolutely but only up to a gauge, which must be the same for all directions.

Consider a system containing two phases, 1 and 2, and a dividing surface. The excess energy, $E^{(s)}$, due to the presence of the dividing surface is the (excess) quantity defined through the relation

$$E = E^{(1)} + E^{(2)} + E^{(s)}, \qquad (1)$$

where E is the total energy of the system and $E^{(1)}$ ($E^{(2)}$) is the energy of the equivalent bulk system of phase 1 (2). The definitions for other surface thermodynamic quantities, such as the entropy, $S^{(s)}$, etc. are similar.[6] The surface free energy per unit area is defined to be

$$\gamma = e^{(s)} - Ts^{(s)} - \sum_i \mu_i \Gamma_i, \qquad (2)$$

where small letters denote the corresponding excess quantities per unit area with the exception of $\Gamma_i = N_i^{(s)}/A$, which is the excess number of particles of the ith species per unit area, and μ_i is the chemical potential of the ith species.

The surface free energy defined in this way measures the work required to create a new surface. It is, in general, distinct from the work needed to deform a surface which is the surface stress. In fluids, as is well-known, the surface free energy is isotropic and is the same as the surface stress.[6] In contrast, the surface free energy in solids is generally anisotropic, which is to say that the surface free energy, $\gamma(\hat{n})$, is a function of the direction of the outward-pointing normal. Further, the surface free energy and the surface stress are distinct quantities.

The surface free energy per unit area is important in the determination of the equilibrium geometry of small crystals and the equilibrium shape of small particles in contact with a substrate. Under conditions of constant temperature T, volume V and chemical potentials μ_i, the excess free energy of the system due to the presence of the surface is

$$\Psi = \int dA \, \gamma(\hat{n}), \qquad (3)$$

where dA is the element of surface area with outward-pointing normal $\hat{n}$. The equilibrium shape thus minimizes the excess free energy.

For a single crystal at constant temperature and chemical potential, the minimization of Eq. (3) subject to the constraint of fixed volume leads to the Wulff construction for determining the equilibrium shape,[6] while a modification of the Wulff construction, the Winterbottom construction,[7] gives the shape of a crystal for a given orientation of the substrate and, thus, information about the growth mode.

Since the surface energy manifests itself physically through the minimization of the integral in Eq. (3), it is possible to add to the surface energy a gauge term $\mathbf{C} \cdot \hat{n}$ (where $\mathbf{C}$ is a constant vector) without changing the equilibrium shape.[8,9] Symmetry, however, permits a nonzero $\mathbf{C}$ for only 10 point groups, namely C_6, C_{6v} (which includes wurtzite), C_4, C_{4v}, C_3, C_{3v}, C_2, C_{2v}, C_{1h} and C_1. For C_1, $\mathbf{C}$ is completely arbitrary; for the rest, symmetry constrains the choice of $\mathbf{C}$. For C_{1h}, $\mathbf{C}$ must lie on the mirror plane; for the remaining eight point groups, $\mathbf{C}$ must lie along the unique axis. In other words, one, two or three arbitrary scalars must be chosen.[9]

II. DETERMINATION OF THE SURFACE ENERGY FROM TOTAL ENERGY CALCULATIONS

For the surface energy two situations exist: (i) for high symmetry systems, the surface energy is unambiguously defined (combining cases one and two of the introduction); (ii) for low symmetry systems, the surface energy is defined up to a gauge. In this section, we begin by pointing out that slab calculations permit a determination of the absolute energy only in high-symmetry cases. In other cases, slab calculations provide only relative surface energies, even if the surface energy is absolutely defined. We then present a method that permits the calculation of absolute surface energies using polyhedra and scaling behavior.

The calculations employed to determine the surface energy for a given crystal direction are generally slab calculations. The slab consists of a finite number of layers and is made infinite in the plane of the surface through the imposition of periodic boundary conditions. The need to decouple the two surfaces from one another dictates the thickness of the slab and also the number of vacuum layers (the calculations are usually carried out using supercells). We define the energy of the equivalent bulk as the sum of the number of atoms of each species times the chemical potential of that species. Invariably, a *single* configuration is used to determine the total *ground state* energy of the slab at T=0. In a multicomponent system, the chemical potential for a single species is not defined for a single configuration. In fact, the chemical potentials that occur in the slab calculations are *external parameters* and can take arbitrary values; however,

physical arguments are invoked to establish limits upon the variability of the chemical potentials of the individual species.

At equilibrium, the chemical potential of a species is equal in all phases that are in contact. This observation can be exploited to impose constraints on the possible equilibrium values. In particular, it is generally assumed that the bulk is in equilibrium with the surface, *i.e.*, that the sum of the chemical potentials of the individual species, μ_A, μ_B, etc. equals the bulk chemical potential $\mu_{AB\cdots Z(bulk)}$,

$$\mu_{AB\cdots Z(bulk)} = \mu_A + \mu_B + \ldots + \mu_Z. \quad (4)$$

Further, the chemical potential for a given species cannot be above the chemical potential of its elemental bulk phase, since the bulk phase would then be unstable with respect to precipitation of the elemental bulk. Thus, the maximum value of the chemical potential of a given species, say A, is equal to the chemical potential of its elemental bulk phase $\mu_{A(bulk)}$. The heat of formation, ΔH_f, which relates the chemical potential in the elemental bulk system to the chemical potential in the compound is defined by

$$\mu_{AB\cdots Z(bulk)} = \mu_{A(bulk)} + \mu_{B(bulk)} + \ldots + \mu_{Z(bulk)} - \Delta H_f. \quad (5)$$

It can be used to set bounds for the chemical potential of the individual species in the bulk, viz.

$$\mu_{\alpha(bulk)} - \Delta H_f \leq \mu_\alpha \leq \mu_{\alpha(bulk)} \quad (6)$$

where $\alpha = A, B, \ldots, Z$.[10] It is worth emphasizing that these bounds only delineate the range over which the bulk is stable. If the chemical potential falls outside of these bounds, the bulk will become metastable and kinetics will determine the subsequent behavior of the system, *i.e.*, whether a change of phase will occur, whether the metastable phase will be very long lived, etc. Thus the bounds Eq. (6) provide a *reasonable* range over which to consider the surface energy.[11]

For certain directions of sufficiently high symmetry, the calculation of the surface energy is simplified. Along these special directions, it is possible to find a slab which has identical surfaces. More explicitly, the surfaces are related by a reflection about the center of the slab and, possibly, by a rotation in the plane of the surface. For a slab of material $AB\cdots Z$ consisting of N_A atoms of species A, N_B atoms of species B, etc., in a sufficiently simple structure (the (001) surface is an example of such a surface); the surface energy is

$$\gamma = \frac{1}{2}(E_{slab} - N_A\mu_A - N_B\mu_B - \ldots N_Z\mu_Z), \quad (7)$$

where E_{slab} is the total energy of the slab.[12] Along lower symmetry directions, the slab calculations give only the surface energy relative to some reference, usually taken to be the 1×1 unreconstructed, unrelaxed surface.

It is possible, at least in principle, to determine the absolute surface energies for the symmetry directions for which slab calculations yield only relative surface energies provided the crystal has a space group of sufficiently high symmetry. Construct a polyhedron with identical faces: the total energy of a polyhedron, E_{poly}, that has n_f identical faces, n_e edges, n_v vertices and contains N_A atoms of species A, N_B atoms of species B, etc., is

$$E_{poly} = N_A\mu_A + N_B\mu_B \ldots + N_z\mu_Z + \alpha_f n_f E_{surf} L^2 \quad (8)$$
$$+ \alpha_e n_e E_{edge} L + \alpha_v n_v E_{vertex},$$

where E_{bulk} is the energy per unit volume of the bulk material comprising the polyhedron, E_{surf} is the surface energy per unit area, E_{edge} is the edge energy per unit length, E_{vertex} is the energy of a vertex and L is the characteristic size of the polyhedron. The α coefficients depend upon the detailed shapes of the surfaces, edges and vertices but the different scaling behavior of each of the energy terms can be exploited to calculate the surface energy. After subtracting off the bulk energy, the remainder consists of surface terms, which vary as L^2, edge terms, which vary as L, and vertex terms, which are independent of L. For a sufficiently large polyhedron, the contribution of the edge terms to the remainder will be negligible and the surface energy can be calculated to order $1/L$, viz.

$$E_{surf} = \frac{1}{\alpha_f n L^2}(E_{poly} - N_A\mu_A - N_B\mu_B \ldots - N_z\mu_Z) \quad (9)$$
$$+ \mathcal{O}(1/L).$$

A two-component system with a zinc-blende lattice provides an example of a crystal with sufficiently high symmetry; a tetrahedron can be constructed which has four faces comprised of (111) surfaces terminated by the same species, A or B.[13] Clearly this approach is valid only for a limited set of symmetry directions. For the 11 Laue point groups and 11 of the 21 remaining point groups that have either two rotation axes or one rotation axis with a mirror plane not through the axis, it is possible to find a polyhedron whose faces are composed of crystallographically equivalent surfaces.[9,13]

The surface energies for other directions can be calculated by slicing off one of the vertices; the surface energy for all but one of the faces is known and, in the scaling limit, the unknown surface energy can be determined to $\mathcal{O}(1/L)$. In this case, too, the polyhedron must be sufficiently large to ensure that edge and vertex effects are negligible. Although such calculations are not feasible at present, the procedure described in this section illustrates the existence of unambiguously defined surface energies for crystals with symmetries as described above. For crystals with lower symmetries, an appropriate convention for defining the vector **C**, the same for all surfaces, must be established. For these systems, the surface energies are measured with respect to a single reference surface.

III. THE ENERGY DENSITY FORMALISM

The total energy of a charge-neutral system of electrons with density $\rho(\mathbf{r})$ within the density functional theory is

$$E = T_s[\rho] + V_H[\rho] + E_{XC}[\rho] + V_{ie}[\rho] + E_{ii}(\{\mathbf{R}_I\}), \quad (10)$$

where $T_s[\rho]$ is the Kohn-Sham kinetic-energy functional, $V_H[\rho]$ is the Hartree functional, $E_{XC}[\rho]$ is the exchange-correlation functional, $V_{ie}[\rho]$ is the electron-ion energy and $E_{ii}(\{\mathbf{R}_I\})$ is the electrostatic self-energy of the atoms located at $\{\mathbf{R}_I\}$. Minimization of this functional subject to the constraint that the number of electrons is fixed leads to the Kohn-Sham equations, the solution of which gives the ground state energy and density of the charge-neutral system.[14,15]

The energy density can be defined[3,4] through the relation

$$E = \int_V d^3 r \mathcal{E}(\mathbf{r}). \quad (11)$$

Consistent with this definition of the energy density, any function $f(\mathbf{r})$ that integrates to zero can be added to the energy density. In principle this presents a problem for the energy-density formalism; such an arbitrary function, a gauge function, would make the integrals over subvolumes arbitrary. In practice, however, such a problem *does not arise*. The energy functional and the energy density are *constructed with each term having a physical motivation* and thus such an arbitrary, unphysical function can be *excluded by construction*. Notwithstanding this exclusion of arbitrary unphysical gauge terms, there is an intrinsic variability of the energy density in any multicomponent system, which is due to the nature of the lattice sum for the Coulomb energy of the ions. This intrinsic variability can be used to incorporate the variability of the externally imposed chemical potential.

The forms of many of the components of the energy density follow in a straightforward manner from the definition, Eq. (11), although their concrete realization depends upon the details of the implementation of the calculation. In particular, the present calculations use a supercell geometry, the pseudopotential formalism and a plane-wave basis. The kinetic-energy density is

$$\mathcal{T}(\mathbf{r}) = \frac{1}{2} \sum_n f_n \nabla \psi_n^*(\mathbf{r}) \cdot \nabla \psi_n(\mathbf{r}), \quad (12)$$

where the $\{\psi_n\}$ are the Kohn-Sham wavefunctions and $\{f_n\}$ their occupations. This symmetric form of the kinetic energy is the more basic form that enters into the variational formulation of quantum mechanics.[16,17] In systems possessing periodic boundary-conditions, the minimization of this functional gives the standard Laplacian-form of the kinetic energy in the Kohn-Sham equations.[18] The use of Bloch's theorem permits the expansion of the Kohn-Sham wavefunctions, viz.

$$\mathcal{T}(\mathbf{r}) = \frac{1}{2} \sum_{i,\mathbf{k}} f_{i\mathbf{k}} \nabla \psi_{i,\mathbf{k}}^*(\mathbf{r}) \cdot \nabla \psi_{i,\mathbf{k}}(\mathbf{r}), \quad (13)$$

where the sum on $\mathbf{k}$ is over the first Brillouin zone. Specializing to a plane-wave basis, it is more efficient, from a computational point-of-view, to evaluate the gradient of the wave-function in reciprocal space and then to Fourier transform into real space, where $\mathcal{T}(\mathbf{r})$ is a point-wise product. In real space, the exchange-correlation energy density is the point-wise product of the electron density and the exchange-correlation energy per electron of the homogeneous, interacting electron gas taken at the local density, *i.e.*,

$$\mathcal{E}_{XC}(\mathbf{r}) = \rho(\mathbf{r})\varepsilon_{XC}(\mathbf{r}). \quad (14)$$

The determination of the energy density for the electrostatic terms requires a careful treatment. The reasons are computational and mathematical. The Hartree potential can be most efficiently calculated for a charge density which has zero net charge, while the lattice sums of the Coulomb potentials of the ions must be regularized through the use of techniques related to those introduced by Ewald. The square of the electric field, which is the negative of the gradient of the Hartree potential, *i.e.*,

$$\mathbf{E}(\mathbf{r}) = -\nabla v_h(\mathbf{r}), \quad (15)$$

gives the Maxwell energy-density, viz.

$$\mathcal{E}_M(\mathbf{r}) = \frac{1}{8\pi}|\mathbf{E}(\mathbf{r})|^2. \quad (16)$$

This term can be most efficiently calculated in a manner similar to the kinetic-energy density. The electric field is first evaluated in reciprocal space and then transformed into real space, where the energy density is calculated as a point-wise product.

The pseudo-ion-pseudo-ion energy per supercell must be much more extensively reworked analytically and the sum restricted to the supercell by use of the minimum-imaging convention in order to obtain

$$E_{ion} = \frac{1}{2} \sum_{I,J \in \Omega}{}' \frac{Z_I Z_J}{R_{IJ}} \mathrm{erfc}\left[\frac{R_{IJ}}{\sqrt{R_{c,I}^2 + R_{c,J}^2}}\right] \quad (17)$$

$$- \frac{1}{2} \sum_{I \in \Omega} \frac{2Z_I}{\sqrt{2\pi} R_{c,I}}.$$

The first term in the pseudo-ion-pseudo-ion electrostatic self-energy is due to the Coulomb interaction between a gaussian pseudo-ion at $\mathbf{R}_I$ and one at $\mathbf{R}_J$, while the second term is the electrostatic self-energy of the charge distribution of the pseudo-ion at $\mathbf{R}_I$. The energy density for this contribution to the total energy is obtained by taking the energy corresponding to an atom at $\mathbf{R}_I$ to be the coefficient of a δ-function centered on that atom, viz.

$$\varepsilon_{ion}(\mathbf{r}) = {\sum_{I,J\in\Omega}}' \frac{Z_I Z_J}{R_{IJ}} \mathrm{erfc}\left[\frac{R_{IJ}}{\sqrt{R_{c,I}^2 + R_{c,J}^2}}\right] \qquad (18)$$

$$\times \alpha_{IJ}\delta(\mathbf{r} - \mathbf{R}_I) - \frac{1}{2}\sum_{I\in\Omega}\frac{2Z_I}{\sqrt{2\pi}R_{c,I}}\delta(\mathbf{r} - \mathbf{R}_I)$$

where the prime on the sum indicates that $I \neq J$. The unknown coefficients α_{IJ} reflect an ambiguity in the resolution of this term into a density; the electrostatic energy between a pseudo-ion at $\mathbf{R}_I$ and one at $\mathbf{R}_J$ is a sum evaluated at the ion positions, not an integral evaluated over all space. When the two ions in the sum are distinct, there is no *a priori* way of assigning a portion of the weights to the contribution of each ion to the sum. The α_{IJ} may be chosen to be different for different pairs of species subject only to the constraint

$$\alpha_{IJ} + \alpha_{JI} = 1, \qquad \text{for } I \neq J. \qquad (19)$$

Thus in a two-component system, there are two additional degrees of freedom, α_{IJ} and α_{JI}, and one constraint so that, in effect, there is one degree of freedom. For convenience, the $\{\alpha_{IJ}\}$ were set equal to one-half in the calculations.

The local-pseudopotential energy can be written as

$$E_{lps} = \sum_{I\in\Omega}\int_\Omega d^3r \rho(\mathbf{r}) U_{sr,I}(\mathbf{r}), \qquad (20)$$

where

$$U_{sr,I}(\mathbf{r}) = U_{lps,I}(\mathbf{r}) - \int_V d^3x \frac{n_{g,I}(\mathbf{x})}{|\mathbf{r} - \mathbf{x}|}, \qquad (21)$$

so that the energy density for the local-pseudopotential is

$$\mathcal{E}_{lps}(\mathbf{r}) = \rho(\mathbf{r}) U_{sr}(\mathbf{r}). \qquad (22)$$

In Eq. (21), $n_{g,I}(\mathbf{x})$ is a Gaussian charge density with integrated charge Z_I and width $R_{c,I}$ centered at $\mathbf{R}_I$.

The non-local pseudopotential is short-ranged in real space. Following Chetty and Martin,[3] the non-local pseudopotential-energy associated with an ion at $\mathbf{R}_I$ is gathered into a delta-function centered on that ion. The smallest volumes that will be considered are much larger than the region over which the non-local pseudopotential is non-zero so that no spurious results are introduced by this treatment of the non-local pseudopotential. The energy density for the non-local pseudopotential is

$$\mathcal{E}_{nlps}(\mathbf{r}) = \sum_{I\in\Omega}\mathcal{E}_{nlps,I}\delta(\mathbf{r} - \mathbf{R}_I), \qquad (23)$$

where the coefficient of the delta-function is

$$\mathcal{E}_{nlps,I} = \sum_{i,\mathbf{k}} f_{i\mathbf{k}}\int_\Omega d^3r \int_\Omega d^3r' \psi_{i,\mathbf{k}}^*(\mathbf{r})\hat{U}_{nlps,I}(\mathbf{r},\mathbf{r}')\psi_{i,\mathbf{k}}(\mathbf{r}').$$

$$(24)$$

Thus $\mathcal{E}(\mathbf{r})$ is a sum of densities, viz.

$$\mathcal{E}(\mathbf{r}) = \mathcal{E}_{kin}(\mathbf{r}) + \mathcal{E}_{XC}(\mathbf{r}) + \mathcal{E}_M(\mathbf{r}) + \mathcal{E}_{ps}(\mathbf{r}) + \mathcal{E}_{ion}(\mathbf{r})$$

$$(25)$$

where $\mathcal{E}_{kin}(\mathbf{r})$ is the kinetic-energy density, $\mathcal{E}_{XC}(\mathbf{r})$ is the exchange-correlation energy density, $\mathcal{E}_M(\mathbf{r})$ is the Maxwell energy density (a particular form of the energy density for the Hartree term), $\mathcal{E}_{ps}(\mathbf{r})$ is the pseudopotential energy density, and $\mathcal{E}_{ion}(\mathbf{r})$ is the ion-ion electrostatic self-energy density.

IV. EVALUATION OF THE SURFACE ENERGY

The present approach to the calculation of the surface energy necessarily involves two parts: the energy density methodology[3,4] and an integration procedure. This approach extends the method developed by Chetty and Martin[3–5] by generalizing the integration method so that it can be used even in systems of low symmetry.

A. The Surface Energy Using Symmetry-Adapted Cells

The approach of Chetty and Martin[3–5] makes use of symmetry-adapted unit-cells for the integration. The boundaries of a symmetry-adapted unit-cell are symmetry planes of the crystal. Therefore, the integral of the energy density over this unit cell is gauge independent and the number of atoms in the cell can be unambiguously determined. The surface energy is then

$$\sigma = \int_{V_s} d^3r \mathcal{E}_{slab}(\mathbf{r}) - \sum_i N_i\mu_i, \qquad (26)$$

where V_s is the volume of the symmetry-adapted unit-cell and N_i is the number of atoms of the ith species inside the cell with $i = A, B, \ldots, Z$. The externally imposed chemical potential of the ith species is μ_i; their sum is subject to the constraint, Eq. (4). This approach can be employed to calculate the surface energy of those directions in which the surface is cut obliquely by these symmetry planes. Chetty and Martin[4] applied their method to the (100) and (111) surfaces of GaAs. The (111) surfaces of a zinc-blende crystal are cut obliquely by the (100) and the (110) planes. However, this approach cannot be used for a crystal whose point group is one of the 10 point groups for which the absolute surface energy is not defined, *i.e.*, there will be surfaces which are not cut obliquely by a sufficient number of symmetry planes. Wurtzite (0001) is an example of such a surface; as the result of the hexagonal symmetry, the appropriate symmetry-adapted cell cannot be defined.

B. The Voronoi Polyhedron

The integration procedure must permit a *consistent* evaluation of the surface energy for those low symmetry systems for which the absolute surface energy cannot be defined. Such an approach is needed because, in general, it is necessary to treat the two surfaces of the slab in supercell calculations differently. For instance, to minimize the charge transfer between surfaces, one of the surfaces may need to passivated and consequently it is desirable to have a methodology that permits the surface energy of each face to be calculated separately. Further, in order to make meaningful *ab initio* predictions about equilibrium crystal shape and preferred growth directions, a database of surface energies of reconstructed surfaces for different directions must be made. This requires that the evaluations of the surface energies be consistent between calculations.

In the approach that we have developed, we view the bulk and the slab as built of blocks. These blocks are constructed according to a well-defined rule, which is a natural generalization of the definition of the familiar Wigner-Seitz cell. The same rule is used to obtain the Voronoi polyhedra in the slab: deep in the slab, where the bulk is recovered, the Voronoi polyhedra are identical to those of the bulk; near the surface, they are deformed. With each block we associate an energy, namely the integral of the energy density over that volume. The "total cost" of constructing a slab and the equivalent bulk is determined by simply adding up the respective energies associated with all of the Voronoi polyhedra. The surface energy is then simply the difference in "cost" between the portion of the slab containing one surface and the equivalent bulk.

In a one-component crystalline system condensed on a lattice without a basis, the integral of the energy density over any volume that is charge neutral is unique; it is the total energy per particle. The arbitrariness of the choice of volume notwithstanding, there is a geometrically motivated volume, namely the Wigner-Seitz cell. The Wigner-Seitz cell possesses the symmetry of the lattice, is space filling and charge neutral. If the crystal has a basis of two identical atoms, then there is a difference between the positions of the atoms making up the crystal and the lattice points. The Wigner-Seitz cell for this crystal is the set of all points closer to a given lattice point than to all other lattice points. As for the case of a lattice without a basis, the Wigner-Seitz cell for the lattice with a two-atom basis is charge neutral, possesses the symmetry of the point group of the lattice and is space filling. However, it contains two atoms. The Wigner-Seitz cell can be generalized to a Voronoi polyhedron about each atom. This polyhedron, which is space filling and charge neutral, is invariant under the largest point subgroup of the space group of the lattice.

Consider a lattice that has a basis of two distinct atoms. As in the case of a lattice with a basis of two identical atoms, the Wigner-Seitz-like volume for a given

atom will have the symmetry of the largest point group which is a subgroup of the space group of the crystal. The union of such volumes with the similarly defined cells for the other atom of the basis will be space filling, but the volume need not be charge neutral. We will generally choose to use a Voronoi polyhedron that has these properties and, because it is most intuitive, is charge neutral. This Voronoi polyhedron is obtained if the faces are translated inward or outward along its normal until the volume is charge neutral. If the normal to a face lies along the vector joining two identical atoms, the face will not be shifted. Only a face that has its normal pointing towards an atom of a different species than the one at the center of the cell will shift. However, it is also possible to use non-charge-neutral polyhedra. Indeed, in strongly ionic systems where the charge exchange between the cation and anion is significant or in C_{60} where the carbon atoms are not identical, it is necessary to use charged Voronoi polyhedra. Because the role of the Voronoi polyhedra is to provide a way of counting atoms and the energies (chemical potentials) associated with them, it is not necessary to add terms to the Hamiltonian when using charged Voronoi polyhedra. Mathematical aspects of the definition of the Voronoi polyhedron are discussed in Appendix I.

The integral of the energy density over the Voronoi polyhedron, Ω_{vp}, in bulk is

$$E^* = \int_{\Omega_{vp}} d^3 r \mathcal{E}(\mathbf{r}) \qquad (27)$$

and will be referred to as the bulk-atom energy. In a one component system, E^* is equal to the chemical potential. The integral of the energy density over the Voronoi polyhedron for species A will be referred to as the bulk-atom energy of species A, etc.

For a multi-component system, the bulk-atom energy satisfies the constraint of equilibrium

$$\mu_{AB...Z} = E_A^* + E_B^* + ... + E_Z^* \qquad (28)$$

where $\mu_{AB...Z}$ is the chemical potential. The total energy of a system of N_A atoms of species A and N_B atoms of species B, etc., is

$$E_{tot} = N_A E_A^* + N_B E_B^* + ... + N_Z E_Z^*. \qquad (29)$$

The bulk-atom energy plays a role identical to the imposed chemical potential in a slab calculation, namely of an externally imposed parameter (cf. Sect. II). It can assume arbitrary values in the bulk system subject only to the constraint of Eq. (28). The same physical constraints that are used to delineate the possible values of the chemical potential also delimit the range of the bulk-atom energy.

For atoms deep inside of a slab where the bulk has been recovered, the Voronoi polyhedra of the slab are identical to the bulk Voronoi polyhedra. The shape of each Voronoi polyhedron depends upon the locations of

the neighboring atoms. Consequently as the surface is approached and atoms are displaced from their bulk positions, the volumes will deform. Because the atoms on the surface do not have any neighbors in the vacuum, the volumes for atoms at the surface will extend out to infinity, cf. the definition in Appendix I and Fig. 1.

As an example, consider diamond. In a diamond crystal, which has space group Fd3m, the Voronoi polyhedron is a snub-tetrahedron[19,20] that has sixteen vertices and sixteen faces. The snub-tetrahedron is invariant under the point group T_d and is pristine, i.e., it is space-filling only if it has ideal proportions.[20] Its four hexagonal faces lie between the nearest neighbors and the atom at the origin, while each of its twelve isosceles triangular faces is a bisecting plane between the atom at the center and a next-nearest neighbor. The zinc-blende crystal structure for a compound semiconductor AB has space group F$\bar{4}$3m. In all but the exceptional case of equal weights for the two species, the Voronoi polyhedra will be deformed snub-tetrahedra, i.e., snub tetrahedra with non-ideal proportions (the union of the two deformed snub-tetrahedra – one for A and the other for B – is space filling). In GaAs, the snub-tetrahedra are very close to pristine.

C. The Surface Energy using Voronoi Polyhedra

The use of Voronoi polyhedra permits the determination of the deviation from bulk behavior in a straightforward manner. For each atom, there is one polyhedron; deep in the slab, the Voronoi polyhedra of the bulk are recovered. A section of the slab that extends from a point deep in the slab out to a point deep in the vacuum and contains N_A atoms of species A, N_B atoms of species B, etc., has a total energy equal to the sum of the energy of each of the Voronoi polyhedra contained in that portion of the slab, i.e.,

$$E_{sec} = \sum_I \int_{\Omega_I} d^3r \mathcal{E}_{slab}(\mathbf{r}), \qquad (30)$$

where Ω_I is the Voronoi polyhedron for an atom at $\mathbf{R}_I$ and I runs over all atoms in the section of the slab. In order to obtain the surface energy, the atoms of the slab can be put into a one-to-one correspondence with the atoms of the bulk. This equivalent bulk is comprised of the appropriate bulk, Voronoi polyhedra. Its total energy is simply

$$E_{bulk} = \sum_I \int_{\Omega_{VP,I}} d^3r \mathcal{E}_{bulk}(\mathbf{r})$$
$$= N_A E_A^* + N_B E_B^* + \ldots + N_Z E_Z^*, \qquad (31)$$

where $\Omega_{VP,I}$ is the Voronoi polyhedron for an atom at $\mathbf{R}_I$ and I runs over all atoms in the equivalent bulk (cf. Eq. (29)). The surface energy is thus

$$E_{surf}(E_I^*) = E_{sec} - E_{bulk} \qquad (32)$$
$$= \sum_I \left[\int_{\Omega_I} d^3r \mathcal{E}_{slab}(\mathbf{r}) - E_I^* \right], \qquad (33)$$

where the sum on I runs over all atoms in the section of the slab.

D. Integrals over the Voronoi Polyhedra

The charge density and the energy density, defined on a grid, must be integrated over the convex Voronoi polyhedra. Due to the role played by Brillouin-zone integrations in electronic-structure calculations, there has been constant interest in the development of techniques for the evaluation of Brillouin-zone integrations.[21-24] The integrals of the energy and charge density over the Voronoi polyhedra are the real-space equivalent to the Brillouin-zone integration.

A careful treatment of contributions to the integral from the boundary is essential to the accurate numerical evaluation of integrals over polyhedra such as occur in the present case. The approach that was used was to decompose the polyhedron into tetrahedra by a Delaunay triangulation.[25] This, however, requires vertices on the surface and edges of the polyhedron. These vertices were obtained by intersecting the lines and planes of the FFT grid with the surface and edges of the polyhedron, respectively. A Delaunay triangulation of this lattice, comprised of the FFT grid and the vertices on the faces and edges, was carried out using GEOMPACK.[26] The values of the functions on the surface of the polyhedron and at the center-of-gravity of the tetrahedra were calculated using a tricubic interpolation. The integration rule for the tetrahedron was of third order.[27] In this way sufficient integration accuracy for the charge and energy in a Voronoi polyhedron was obtained.

V. APPLICATIONS

The present approach was used to calculate the surface energies of the (001) and (111) unreconstructed, unrelaxed surfaces of GaAs and of the unreconstructed, unrelaxed surfaces and a few reconstructions of the (0001) surfaces of wurtzite SiC and GaN. In each case, the energy-density formalism with Voronoi polyhedra was used to calculate the reference surface energy of the unrelaxed, unreconstructed surfaces.

The Car-Parrinello method[28-30] was employed with a plane-wave cut-off of 14 Ry for GaAs and 30 Ry for SiC and GaN. The Perdew-Zunger parameterization of the exchange-correlation energy[31,32] was chosen. Norm-conserving pseudopotentials[33-35] were used for gallium, arsenic, carbon, hydrogen and silicon, while the nitrogen pseudopotential was an optimized soft-core pseudopotential of Li and Rabii.[36] The pseudopotentials were p-local

pseudopotentials, except for the gallium pseudopotential which was d-local. The non-local potentials were employed using the Kleinman-Bylander approach.[37] For GaAs, the gallium pseudopotential included the nonlinear core-correction.[38,39]

A supercell was used to reimpose the translational invariance that is broken along the direction parallel to the surface normal. For the calculations involving the (001) gallium-terminated surface, the 2×2 supercell contained 6.5 bilayers of GaAs (both faces gallium-terminated) and 2.5 bilayers of vacuum; for the calculations of the (111) surfaces, the 2×2 supercell contained 6 bilayers of GaAs and three bilayers of vacuum. A 2×4 supercell containing five bilayers of material and 3 bilayers of vacuum was used for SiC; the 2×2 GaN supercell contained six bilayers of material and four bilayers of vacuum.

In order to minimize charge transfer between the faces, hydrogen atoms passivated one face of the GaN and SiC slabs. This was successful as charge transfer was found to be small.

If the work functions of the two surfaces are different, then the periodic boundary conditions of the supercell enforce a common electrostatic potential in the vacuum and result in an unphysical change in the electric potential in the vacuum region equal to the difference between the two work functions. The field, induced by this change, was cancelled by adding a dipole layer in the vacuum region.

A. Surface Energy of GaAs (001)

The calculation of the surface energy of the (001) surface of GaAs provides a good test of the present formalism. In this case a slab in which both surfaces have the same termination exists. The surface energy can thus be calculated using the total energy approach and the surface energies of the two surfaces can be independently calculated using the energy-density approach presented herein. Fig. 2 shows the gallium surface energy calculated with the total energy method compared to that obtained with the energy-density formalism with Voronoi polyhedra for three different cases: 1) charge-neutral Voronoi polyhedra; 2) charged Voronoi polyhedra in which the polyhedra around the gallium atoms have charge 2.85 e, while those centered on the arsenic atoms have charge 5.15 e; 3) charged Voronoi polyhedra in which the polyhedra centered on the gallium atoms have charge 3.15 e, and the ones about the arsenic atoms have charge 4.85 e. Only the calculations of one face of the slab are shown. Similar results are obtained for the other face.

The surface energies agree to within less than 0.1 eV over the entire range. The use of a finer mesh will improve the agreement. In all of the cases, the linear dependence of the surface energy upon the gallium chemical potential has a slope of -0.50, which is in very good agreement with the exact value of -0.50. The value of this slope is not assumed in the calculation, but is a result which confirms the correctness of the approach. Thus it is possible to use either neutral or charged Voronoi polyhedra.

B. Surface Energy of GaAs (111)

The surface energies for the (111) Ga- and As-terminated faces are shown in Fig 3. As expected, a linear dependence upon the gallium chemical potential is observed with a slope of -0.27 for the gallium-terminated surface and 0.26 for the arsenic-terminated surface. This should be compared with the analytical results of -0.25 for the Ga-terminated surface and 0.25 for the As-terminated surface. The energy of the As-terminated face is lower than that of the Ga-terminated face over the entire range and the numerical results are in good agreement with those of Chetty and Martin.

C. SiC surface energy

The surface energies for selected reconstructions[40] of the Si- and C-terminated faces are shown in Fig. 4. The unrelaxed, unreconstructed surfaces are observed to depend linearly on the chemical potential of silicon. In both cases the 2×2 π-bonded reconstruction has the lowest energy of the reconstructions considered. It should be noted, however, that lower energy reconstructions exist and are currently being studied.

D. GaN surface energy

The surface energies for the Ga- and N-terminated faces are shown in Fig. 5. With Γ-point sampling we find that on the gallium-terminated surface the nitrogen-adatom and gallium-vacancy structures have the lowest energies over the entire physical range; the energy difference is smaller than the accuracy of the calculation. In the nitrogen-adatom reconstruction, the three gallium atoms bonded to the nitrogen adatom move towards the adatom. The remaining gallium atom relaxes inwards towards the nitrogen layer. The subsurface N atom, which has three bonds with the surface gallium atoms bonded to the N-adatom, moves outwards towards the Ga atoms, becoming almost coplanar with them. In the vacancy reconstruction, the three gallium atoms relax inwards towards the subsurface nitrogen layer. On the nitrogen-terminated surface, the nitrogen vacancy has the lowest energy while the gallium-adatom has a somewhat higher energy. Although the energies of the individual surfaces can be shifted by a constant (cf. Section I), the sum of the surface energies is an absolutely defined quantity.

VI. SUMMARY AND CONCLUSIONS

The surface energy plays a critical role in determining equilibrium crystal shapes and modes of growth. The standard *ab initio* method for the evaluation of surface energies employs total energy slab calculations. These calculations give the absolute surface energies only for those exceptional directions in which the two faces of the slab are crystallographically the same. For other surfaces, they give the energies relative to some reference surface, even in those cases for which the absolute surface energy exists. In order to evaluate the surface energies of lower symmetry surfaces (for which the absolute surface energy is defined) Chetty and Martin[3-5] introduced the energy density formalism together with symmetry-adapted unit-cells; however, energy-density approach procedure is not applicable to systems where such cells cannot be defined.

The surface energy, which is an excess thermodynamic quantity, is absolutely defined for crystals which have a center of symmetry, or two axes of rotation or an axis of rotation and a mirror plane not through this axis. For lower symmetry crystals, the surface energy is defined up to a gauge term $\mathbf{C} \cdot \hat{\mathbf{n}}$, where $\mathbf{C}$ is constrained by symmetry for all but the point group C_1; hence for these systems the surface energy must be defined in a consistent manner for all directions.[8,9]

We have developed a generalization of their approach that permits the evaluation of surface energies in all cases, subject only to the physical ambiguity of the gauge term $\mathbf{C} \cdot \hat{\mathbf{n}}$. Our method uses appropriately defined Voronoi polyhedra to calculate the differences between the bulk and surface contributions to the integrated energy density, thus isolating the excess energy introduced by the surface.

This approach was used to calculate the surface energy of gallium-terminated GaAs (001). Since the surface energy in this case can be evaluated using a total energy calculation, it provides a good test of the method. The surface energies of the two approaches (total-energy and energy-density with Voronoi polyhedra) agreed to less than 0.1 eV per surface atom. The surface energies of the gallium- and arsenic-terminated faces of GaAs (111) were determined and found to be in agreement with those of Chetty and Martin.[4] Finally, the approach was used to evaluate the surface energies of selected reconstructions of the (0001) surfaces of the wide-gap semiconductors SiC and GaN.

APPENDIX I: MATHEMATICAL DETAILS CONCERNING THE VORONOI POLYHEDRON

For a Bravais lattice which has lattice points $\{\mathbf{x}_j\}$, the region of space nearer to the lattice point $\mathbf{x}_i$ than to $\mathbf{x}_j$ (the so-called dominance region of $\mathbf{x}_i$ over $\mathbf{x}_j$) is the half-space

$$D(\mathbf{x}_i, \mathbf{x}_j) = \{\mathbf{x} | \quad |\mathbf{x} - \mathbf{x}_i|^2 < |\mathbf{x} - \mathbf{x}_j|^2\}; \qquad (34)$$

the boundary of this half-space is the plane defined by

$$|\mathbf{x} - \mathbf{x}_i|^2 = |\mathbf{x} - \mathbf{x}_j|^2; \qquad (35)$$

and the Wigner-Seitz cell is the intersection of the all such half-spaces, *i.e.*,

$$V(\mathbf{x}_i) = \cap_{j \neq i} D(\mathbf{x}_i, \mathbf{x}_j). \qquad (36)$$

Since the Wigner-Seitz cell is the intersection of convex half-spaces, it is convex. Because its faces are planes, it is a polyhedron.

The isogonal point group of a crystal is the group formed from all of the point-group operations which occur in the space group.[41] If the space group is symmorphic, the isogonal point group will be a subgroup of the space group; if the space group is non-symmorphic (it contains either a screw axis or glide-reflection plane), it will not be a subgroup. Because the symmetry operations of the point group leave the lattice points unchanged, the Wigner-Seitz cell is also invariant under the operations of the point group. The Wigner-Seitz cell is also space filling and is charge neutral. Thus the three properties that characterize a Wigner-Seitz cell are: it is a space-filling polyhedron that has the symmetry of the point group of the lattice and is charge neutral.

In the case of diamond, there are two carbon atoms in the basis and the Wigner-Seitz cell as traditionally defined contains the same number of atoms as there are atoms in the basis, *i.e.*, two carbon atoms. A Wigner-Seitz cell for an atom or, more properly, a Voronoi polyhedron, can defined by analogy with the Wigner-Seitz cell for a lattice point. The region nearer to an atom at $\mathbf{R}_I$ than to an atom at $\mathbf{R}_J$ is the half space

$$D(\mathbf{R}_I, \mathbf{R}_J) = \{\mathbf{x} | \quad |\mathbf{x} - \mathbf{R}_I|^2 < |\mathbf{x} - \mathbf{R}_J|^2\}. \qquad (37)$$

The intersection of the all these half-spaces, *i.e.*,

$$V(\mathbf{R}_I) = \cap_{J \neq I} D(\mathbf{R}_I, \mathbf{R}_J) \qquad (38)$$

defines the Voronoi polyhedron. Any point inside this cell is closer to the atom at $\mathbf{R}_I$ than to all other atoms (rather than to the lattice point as is the case for the traditional Wigner-Seitz cell). The Voronoi polyhedron, being the intersection of half-spaces, is a convex polyhedron.

Defined in this way, the Voronoi polyhedron possesses many properties similar to those of the standard Wigner-Seitz cell. It is a space-filling polyhedron and charge neutral. However, it is no longer necessarily invariant under

the symmetry operations of the isogonal point group, but only under the operations of the largest point group that is a subgroup of the space group of the crystal. Further, the integral of the energy density over this volume is independent of any gauge and is equal to the total energy per atom.

If the basis atoms are not identical, for example as in GaAs, the Voronoi polyhedron will no longer be neutral as is desired. In this case, the Voronoi polyhedra is defined to be an additively-weighted power Voronoi polyhedron.[25] The additively-weighted power Voronoi polyhedron or, more simply, the power Voronoi polyhedron extends *naturally* the definition of the Voronoi polyhedron from a lattice with a basis of identical atoms to a lattice with a basis of distinct atoms. In point of fact, the standard Wigner-Seitz cell, the Voronoi polyhedron for a lattice with a basis of identical atoms and the Voronoi polyhedron for a lattice with a basis of distinct atoms are all encompassed within the definition of the power Voronoi polyhedron.

The additively-weighted power-distance of point x from an atom at $\mathbf{R}_I$ is

$$d_{pw}(\mathbf{x}, \mathbf{R}_I; w_I) = |\mathbf{x} - \mathbf{R}_I|^2 - w_I \qquad (39)$$

where w_I is the weight associated with the atom at $\mathbf{R}_I$. With respect to the additively-weighted power-distance, the half-space closer to an atom at $\mathbf{R}_I$ than to one at $\mathbf{R}_J$ is

$$D(\mathbf{R}_I, \mathbf{R}_J) = \{\mathbf{x}| \quad |\mathbf{x} - \mathbf{R}_I|^2 - w_I < |\mathbf{x} - \mathbf{R}_J|^2 - w_J\} \qquad (40)$$

and the power Voronoi polyhedron is the intersection of the half-spaces, *i.e.*,

$$V(\mathbf{R}_I) = \cap_{J \neq I} D(\mathbf{R}_I, \mathbf{R}_J). \qquad (41)$$

The use of the additively-weighted power-distance is motivated by the observation that the planes defining the surfaces of the traditional Wigner-Seitz cell are defined through Eq. (35) and that the additive weights rigidly translate the faces of the Wigner-Seitz cell, always maintaining a polyhedral shape. When $w_I = w = constant$ for a lattice without a basis, the standard Wigner-Seitz cell is obtained.

For a tetrahedrally-coordinated compound semiconductor comprised of elements A and B, the volume contained within the Voronoi polyhedron is a monotonic function of the difference of the weights w_A and w_B. If charge-neutral polyhedra are used, the value of this difference is fixed by the requirement that the total electronic charge within the Voronoi polyhedron cell be equal to the ionic charge; if, instead, charged polyhedra are used, the difference can be chosen for convenience. In the case of C_{60}, not all of the atoms are in symmetry equivalent positions and charged Voronoi polyhedra must be used. Further, the space group is nonsymmorphic. Consequently, the largest point group that is a subgroup of

the space group for solid C_{60} is not T_h^6, but $\bar{3}$ and the Voronoi polyhedra will have this symmetry.

---

[1] J. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy* (Academic Press, Boston 1993).

[2] J. LaFemina, Phys. Rep. **16**, 133 (1992).

[3] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6074 (1992).

[4] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6089 (1992).

[5] N. Chetty and R. M. Martin, Phys. Rev. B **44**, 5568 (1991).

[6] C. Herring, in *The Physics of Powder Metallurgy*, ed. W. E. Kingston (Am. Soc. Metals, Cleveland, Ohio, 1951) pp. 7-8.

[7] W. Winterbottom, Acta Metall. **15**, 303 (1967).

[8] J. Lee, H. Aaronson and K. Russel, Surf. Sci. **51**, 302 (1975).

[9] E. Arbel and J. Cahn, Surf. Sci. **51**, 305 (1975).

[10] G-X. Qian, R. Martin and D. Chadi, Phys. Rev. B **38**, 7649 (1988).

[11] There may also be other stoichiometries that will further limit the stability range, thus making this choice somewhat arbitrary, although conventional.

[12] This definition of the surface energy (at T=0) corresponds to Herring's specific surface (free) energy.

[13] M. Buerger, *Elementary Crystallography*, (MIT Press, Cambridge 1978).

[14] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864B (1964).

[15] W. Kohn and L. Sham, Phys. Rev. **140**, 1133A (1965).

[16] J. Slater, Phys. Rev. **51**, 846 (1937).

[17] D. Cook, *Schrödinger's Mechanics*, (World Scientific, Singapore 1988).

[18] There is, in principle, no problem in choosing a different form for the kinetic energy density, provided it is used consistently. The choice serves to change the gauge.

[19] L. Föppl, Phys. Zeitschr. **15**, 191 (1914).

[20] K. Critchlow, *Order in Space: A Design Source Book*, (The Viking Press, New York, 1969).

[21] O. Jepson and O.K. Anderson, Solid State Commun. **9**, 1763 (1971).

[22] G. Lehmann and M. Taut, Phys. Stat. Sol.(b) **54**, 469 (1972).

[23] H. Monkhorst and J. Pack, Phys. Rev. B **13**, 5188 (1976).

[24] P. Blöchl, O. Jepson and O.K. Anderson, Phys. Rev. B **49**, 16223 (1994).

[25] A. Okabe, B. Boots and K. Sugihara, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, (Wiley, Chichester, 1992).

[26] B. Joe, Adv. Eng. Software, **13**, 325 (1991).

[27] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, (Dover, N.Y. 1972), p.895.

[28] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[29] M. Payne, M. Teter, D. Allan, T. Arias and J. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[30] D. Remler and P. Madden, Mol. Phys. **70**, 921 (1990).

[31] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[32] D. Ceperley and B. Alder, Phys. Rev. Lett. **45** 566 (1980).

[33] G. Bachelet, D. Hamann and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[34] D. R. Hamann, M. Schlüter and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[35] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[36] G. Li and S. Rabii, unpublished (1992).

[37] L. Kleinman and D. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[38] S. Louie, S. Froyen and M.L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[39] M. Buongiorno Nardelli, K. Rapcewicz and J. Bernholc, to be published (1996).

[40] Y. Hwang, B. Chen, P. Bogusławski and J. Bernholc, MRS Symp. Proc. **327**, 293 (1994).

[41] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids*, (Oxford, Clarendon Press, 1972).

FIG. 2. Comparison of the surface energy of the gallium-terminated surface of GaAs(001) from a total energy calculation with the values obtained using the energy-density formalism with charge-neutral and charged Voronoi polyhedra (the charge of the Voronoi Polyhedron (VP) centered on the gallium atom in example 1 is 2.85 e, while in example 2 it is 3.15 e). The results agree to less than 0.1 eV. The slope predicted from the formalism is -0.50 in excellent agreement with the theoretical result of -0.50. The range of the abscissa is chosen, by convention, to be twice the heat of formation. In this and the following figures, it has been slightly broadened to include the possibility of metastability as discussed in the text (Sect. II).

FIG. 1. A two-dimensional slab showing the atomic cells. In the center of the slab, the bulk Voronoi polyhedra are recovered; the cells at the surface extend to infinity.
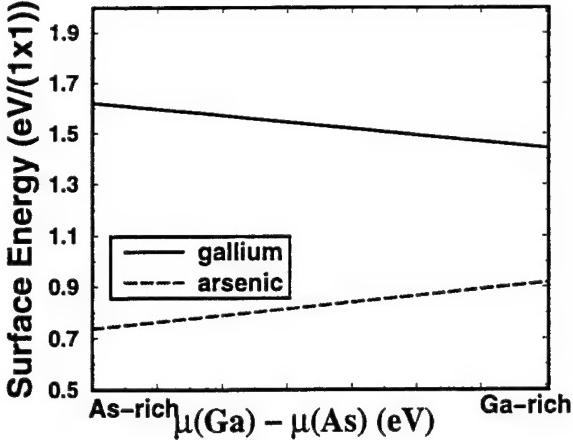
FIG. 3. Surface energy of the unrelaxed, unreconstructed GaAs(111) surfaces calculated using the energy-density formalism with Voronoi polyhedra.
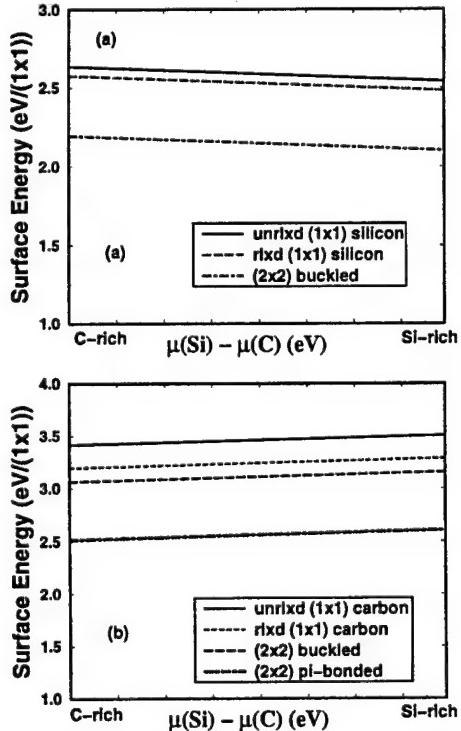

FIG. 4. Surface energy of selected reconstructions of SiC (0001) calculated using the energy-density formalism with Voronoi polyhedra: (a) the silicon-terminated surface and (b) the carbon-terminated surface. The linear dependence of the surface energies of the reconstructions of the carbon-terminated surface on the carbon chemical potential has slope -0.24; the linear dependence of the surface energies of the reconstructions of the silicon-terminated surface on the silicon chemical potential has slope -0.24.
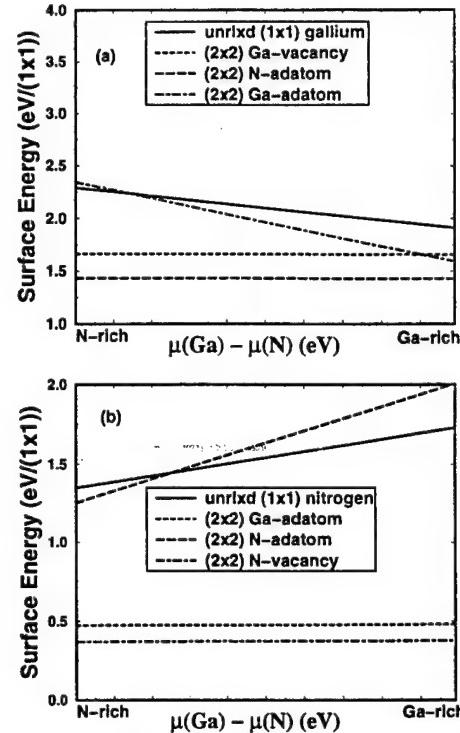
FIG. 5. Surface energy of the reconstructions of GaN (0001) calculated using the energy-density formalism with Voronoi polyhedra: (a) the gallium-terminated surface and (b) the nitrogen-terminated surface. For the unreconstructed, unrelaxed gallium-terminated face, the surface energy depends linearly upon the gallium chemical potential with slope of -0.25; for the unreconstructed, unrelaxed nitrogen-terminated face, the surface energy depends linearly upon the nitrogen chemical potential with slope of -0.25.

V. Theory of Interfaces and Surfaces in Wide Gap Nitrides

Marco Buongiorno Nardelli, Krzysztof Rapcewicz and J. Bernholc

Department of Physics, North Carolina State University, Raleigh, NC 27695-8202

We review a selection of the results of a theoretical investigation of the properties of interfaces and surfaces of the wide-gap III-V nitride semiconductors. The electronic properties of wurtzite heteroepitaxial interfaces of AlN and GaN, incorporating the effects of strain are discussed. In particular, we find that this interface is of Type I and have calculated the valence-band offset to be -0.57 eV. We also present the surface energies and atomic geometries of the 2×2 reconstructions of the (0001) face of GaN. In conditions which are rich in a given species, an adatom reconstruction of that species is found to be the most energetically favorable: for gallium-rich conditions, the reconstruction with a gallium-adatom on a T_3-site is the most stable, while for nitrogen-rich conditions the reconstruction with a nitrogen-adatom on the H_3 site is energetically the most favorable.

Semiconductor heterojunctions find wide applications in microelectronics. These include lasers, photodetectors, high-efficiency solar cells and radiation-resistant integrated circuits.[1] The III-V nitrides with their wide band-gaps, high thermal stability and conductivity and radiation resistance have long been viewed as attractive candidates for use in microelectronic devices.[2]

The recent demonstration of a GaN-based blue laser, which follows upon three decades of research, has served to highlight the potential of these materials.[3] However, at present, the mechanisms underlying the workings of the blue laser are not understood. To this end and in order to develop other nitride-based devices, a great effort is being expended to understand and control the materials properties of the wide-gap nitrides.

METHODOLOGY

The standard *ab initio* plane-wave pseudopotential method[4-6] was used in the calculations of interface properties. To ensure convergence of the nitrogen pseudopotential, an energy cut-off for the plane-wave expansion of 50 Ry was chosen. The equivalent of 10 k-points for bulk and superlattice calculations in the zinc-blende structure[7] and 6 k-points for calculations of the wurtzite structure[8] were employed. A careful check of the convergence in both the size of the plane-wave basis and the number of special points was made. For the surface calculations, a multi-grid-based total-energy method that uses a real-space grid as the basis was used. This approach has been used to study a wide range of systems

involving large numbers of atoms.[9] The ions were relaxed using the "fast-relax" algorithm.[10]

The Perdew-Zunger parametrization[11] of the Ceperley-Alder form[12] of the exchange-correlation energy was chosen. Non-local, norm-conserving pseudopotentials[13-15] were included using the Kleinman-Bylander approach.[16] For gallium, a recently-developed norm-conserving pseudopotential, which includes a non-local core correction,[17] and permits an efficient description without the need for an explicit treatment of the d-valence electrons was used. For nitrogen, a standard pseudopotential with a neutral configuration as the atomic reference for all states was employed. These pseudopotentials have been demonstrated to reproduce accurately the bulk properties of GaN.[19]

The calculated bulk properties are presented in Table I. The theoretical lattice parameters of both the zinc-blende and wurtzite forms agree very well with experiment; a similar level of accuracy is expected for the surface and interface calculations described below. The cohesive energy of GaN and α-Ga is 10.42 eV and 3.4 eV, respectively. The binding energy of nitrogen molecule is 5.87 eV per nitrogen atom, in good agreement with other LDA calculations.[20] The theoretical heat of formation of GaN is thus 1.15 eV, in good agreement with the experimental value of 1.14 eV.

INTERFACE PROPERTIES

Interface behavior is very important in determining the properties of semiconductor devices. We have investigated the wurtzite (0001) GaN/AlN interface. Superlattices of AlN and GaN are expected to be strained because of the 3.5% theoretical lattice-mismatch between AlN and GaN (the experimental value is 2.7%). The effects of the strain were incorporated using macroscopic elasticity theory.[18] In this theory, each half of the heterojunction is treated as a strained bulk with a fixed in-plane lattice constant $a_{\parallel}$. The strain energy of the system is minimized, keeping the in-plane lattice constant fixed, to determine the perpendicular lattice constant, c, of the epilayer. For AlN and GaN, the calculated elastic constants used to determine the strains are given in Table II. Total energy calculations confirm the that macroscopic elasticity theory predicts well the perpendicular lattice constant for the epilayer. The residual relaxation of the atoms at the interface is negligible (< 0.05 Å) and does not affect the band offset of the interface.[19]

The band offsets of the (0001) GaN/AlN strained heterojunction were studied following the procedure of

Ref. 21. The valence band offset is divided into a band-structure contribution, which is the difference between the energies of the valence-band and conduction-band edges when the average electrostatic potentials of the epilayers are aligned, and the difference in the value of the average electrostatic potential (ΔV) in the two epilayers of the heterostructure. Strain will affect the electronic properties of the interface through the variation of the average electrostatic potentials and through the variations of the band edges (deformation potentials). The calculated valence-band offset in the case of an AlN in-plane lattice constant (strained GaN) is estimated to be -0.57 eV,[22] a value smaller than the result for the strained non-polar (001) GaN/AlN interface. The ratio of the conduction-band to valence-band offset is 65:20. These results agree very well with the experimental measurements of the (0001) wurtzite interface.[23]

As spin-orbit effects in both AlN and GaN have been shown to be of the order of 20 meV, their difference, which enters into the calculations of the band offsets, is much smaller than other possible sources of systematic error.[24] These include the neglect of the anion p- and cation d-state repulsion[25] and the well-known neglect of many-body effects in the LDA. In GaN/AlN interfaces, the inclusion of the $3d$-electrons as valence electrons results in a constant shift of 0.2 eV, which is less than the experimental error,[23] and does not change the character of the interface. Incorporating this shift gives results in agreement with previous work using a d-valence pseudopotential[30] and an all-electron calculation.[26] The importance of many-body effects on the band offsets is not known and awaits a future GW calculation.

The wurtzite system displays pyroelectric and piezo-electric behavior.[27] These effects will manifest themselves macroscopically in multiple-quantum-wells along those directions that do not have a perpendicular mirror-plane.[28,29] In (0001) strained GaN/AlN, we observe a substantial electric field as has been previously noted by Satta and coworkers.[30] We have calculated the spontaneous bulk polarization of unstrained AlN and the strain-induced polarization for the GaN epilayer, in order to distinguish the bulk pyroelectric and piezoelectric contributions to this field from that induced by the interface. A superlattice consisting of 8 layers of wurtzite and 6 layers of lattice-matched zinc-blende was used.[31] In unstrained zinc-blende, the bulk spontaneous polarization is rigorously zero. Further, zinc-blende-wurtzite interface does not introduce any chemical or geometrical perturbations. This construction, therefore, permits the unambiguous determination of the spontaneous polarization present in the wurtzite structure from the slope of the macroscopic average of the electrostatic potential. The spontaneous polarization (P_3) of AlN and GaN in equilibrium is -1.227 $\mu C/cm^2$ and -0.448 $\mu C/cm^2$ respectively; the polarization of the strained GaN is -0.454 $\mu C/cm^2$. These values are comparable to the computed bulk polarization in BeO.[31] The effect of the interface dipole is small; the polarization in the GaN/AlN multiple-quantum-well is well described as the sum of the polarizations of the constituent epilayers of the multiple-quantum-well. The estimated contribution of the interface dipole (which includes the response of one epilayer to the field of the other) is 0.057 $\mu C/cm^2$, which is of opposite sign and an order of magnitude smaller than the bulk polarizations. The computed value of the polarization in the superlattice agrees with that estimated from experiment by Martin et al.[23]

SURFACE RECONSTRUCTIONS

The (0001) surface is a polar surface. The ideal surface has dangling bonds with charge distributions that are energetically unfavorable. If kinetically permitted, the surface atoms will relax and redistribute the "dangling-bond" charge density so as to satisfy the valences of all of the surface species. This can be achieved via structural relaxation (with a concomitant charge transfer between surface atoms). Recent experimental studies have revealed the presence of 2×2 reconstructions during growth.[32] In this paper, we restrict our attention to these reconstructions on the gallium-terminated face.

We considered eight different 2×2 reconstructions, namely the ideal relaxed structure and gallium vacancy, nitrogen adatom, gallium adatom and nitrogen-trimer, each on the T_3 and H_3 (hollow) sites (See Figs. 1 and 2). We find that adatom reconstructions are the most energetically favorable of the studied reconstructions. In particular, in the gallium-rich case that the gallium adatom on the T_3 site has the lowest energy, while in the nitrogen-rich case the nitrogen adatom on the H_3 is the most stable.

The slabs contained four bilayers of GaN, three bilayers of which were relaxed. Relaxations in the third bilayer were negligible, indicating that a sufficiently thick slab was employed. Pseudo-hydrogens with $Z=0.75$ passivated the nitrogen-terminated face.[33] If the work functions of the two surfaces are different, then the periodic boundary conditions of the supercell enforce a common electrostatic potential in the vacuum and result in an unphysical change in the electric potential in the vacuum region equal to the difference between the two work functions. A large vacuum region of 10 Å served to reduce the size of this field. Previous calculations that explicitly included a correction for this field did not show significant differences in the final equilibrium geometries,[34] and consequently field corrections were not included in our calculations. An orthorhombic supercell with grid spacing of 0.21, 0.22 and 0.20 a.u. in the x-, y- and z-directions, repectively, was employed. The eight reconstructions mentioned above were studied using Γ-point sampling. For the lowest energy structures of each class i.e., nitrogen adatom on H_3 site, nitrogen trimer on the T_3 site, etc., a further calculation using two k-points in the irreducible Brillioun zone was carried out.

Fig. 3 displays the relative surface energies for the five different reconstructions calculated using 2 k-points in the irreducible Brillouin zone. Under gallium-rich conditions and for most of the range of the chemical potential, the lowest energy reconstruction is the gallium-adatom on the T_3-site; while under nitrogen-rich conditions, the reconstruction with the nitrogen-adatom on the H_3-site is preferred energetically.

The gallium-adatom reconstruction is shown in Fig. 1. The adatom sits on the T_3-site above the subsurface nitrogen atom; the gallium-nitrogen distance is 2.46 Å. The gallium-adatom surface-gallium bond length is 2.40 Å, which is very close to the Ga-dimer distance in α-gallium (2.44 Å). The bond angle between two gallium surface atoms and the adatom is 82°. The in-plane relaxation of the surface atoms is negligible; the only observed relaxation of the first bilayer is along the z-direction. The three gallium atoms bonded to the adatom remain coplanar to a significant degree, while the remaining unbonded gallium surface atom relaxes into the slab. The proximity of the gallium adatom to the nitrogen subsurface atom lowers the electrostatic energy and makes this particular configuration energetically more favorable than the gallium adatom reconstruction with the adatom on the hollow site.

The nitrogen-adatom reconstruction is shown in Fig. 2. The adatom sits on the hollow (H_3) site. The nitrogen-adatom surface-gallium bond-length is 2.0 Å and is very close to the Ga-N bond length distance in GaN bulk (1.96 Å). The bond angle between two gallium surface atoms and the nitrogen adatom is 89°. The relaxation perpendicular to the surface is very similar to that in the gallium-adatom reconstruction. This configuration is stabilized because it separates the nitrogen adatom from the subsurface nitrogen atom and thereby lowers the electrostatic repulsion.

SUMMARY

We have reviewed a selection of theoretical results for the properties of interfaces and surfaces of the gallium and aluminium nitrides. The electronic properties of wurtzite heteroepitaxial interfaces of AlN and GaN, incorporating the effects of strain were discussed. The strained AlN/GaN interface is of Type I and has a valence-band offset of -0.57 eV. As a result of the pyro- and piezoelectric nature of the wurtzite nitrides, macroscopic polarization of 0.057 $\mu C/cm^2$, is present in multiple-quantum-wells along the (0001) direction. The surface energies and atomic geometries of the 2×2 reconstructions of the (0001) face of GaN were also investigated. Adatom reconstructions were found to be the most energetically favorable over the whole range of gallium chemical potential: under gallium-rich conditions, the reconstruction with a gallium-adatom on a T_3-site is the most energetically favorable, while under nitrogen-rich conditions the reconstruction with a nitrogen-adatom on the H_3 site has the lowest energy.

[1] R. Bauer and G. Margaritondo, Physics Today, **40**, 27 (1987).

[2] M.-E. Lin, B. N. Sverdlov, S. Strite, H. Morkoç and A.E. Drakin, Electron. Lett. **29**, 1759 (1993).

[3] S. Nakamura et al., Jpn. J. Appl. Phys. **35**, L74 (1996).

[4] M. Payne, M. Teter, D. Allan, T. Arias and J. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[5] W. Pickett, Comp. Phys. Rep. **9**, 115 (1989).

[6] D. Remler and P. Madden, Mol. Phys. **70**, 921 (1990).

[7] S. Froyen, Phys. Rev. B, **39**, 3168 (1989).

[8] D.J. Chadi and M.L. Cohen, Phys. Rev. B **8**, 5747 (1973).

[9] E. L. Briggs, D. J. Sullivan and J. Bernholc, Phys. Rev. B **52**, R5471 (1995); E. L. Briggs, D. J. Sullivan and J. Bernholc, Phys. Rev. B **54**, 14362 (1996).

[10] C. Wang, Q.-M. Zhang, and J. Bernholc, Phys. Rev. Lett. **69**, 3789 (1992).

[11] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[12] D. Ceperley and B. Alder, Phys. Rev. Lett. **45** 566 (1980).

[13] G. Bachelet, D. Hamann and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[14] D. Hamann, M. Schlüter and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[15] D. Hamann, Phys. Rev. B **40**, 2980 (1989).

[16] L. Kleinman and D. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[17] S. Louie, S. Froyen and M.L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[18] C. Van de Walle and R.M. Martin, Phys. Rev. B **34**, 5621 (1986).

[19] M. Buongiorno Nardelli, K. Rapcewicz and J. Bernholc, Phys. Rev. B **55** R7323 (1997).

[20] B.G. Johnson, P. Gill and J.A Pople, J. Chem. Phys. **98** 5613 (1993).

[21] A. Baldereschi, S. Baroni and R. Resta, Phys. Rev. Lett. **61**, 734 (1988).

[22] The band-offset was calculated following A. Munoz, N. Chetty and R. M. Martin, Phys. Rev. B **41**, 2976 (1990). The position of the interface was unambiguously determined from the position of the interface dipole.

[23] G. Martin et al., Appl. Phys. Lett. **65**, 610 (1994), *ibid.* **68**, 2541 (1996).

[24] M. Suzuki, T. Uenoyama and A. Yanase, Phys. Rev. B **52** 8132 (1995).

[25] S. Wei and A. Zunger, Phys. Rev. Lett. **59**, 144 (1987).

[26] E. Albanesi, W. Lambrecht and B. Segall, J. Vac. Sci. Technol. B **12**, 2470 (1994).

[27] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunder College, Philadelphia 1976). Ch. 27.

[28] D. Smith, Solid State Commun. **57**, 919 (1986).

[29] A. Bykhovski, B. Gelmont and M. Shur, Appl. Phys. Lett. **63**, 2243 (1993); J. Appl. Phys. **74**, 6734 (1993).

[30] A. Satta, V. Fiorentini, A. Bosin, F. Meloni and D. Van-

derbilt, MRS Proceedings **395**, 515 (1996).

[31] M. Posternak, A. Baldereschi, A. Catellani and R. Resta, Phys. Rev. Lett. **64**, 1777 (1990).

[32] M. A. L. Johnson, Z. Yu, C. Boney, W. H. Rowland Jr., W. C. Hughes, J. W. Cook Jr., J. F. Schetzina, N. A. El-Masry, M. T. Leonard, H. S. Kong and J. A. Edmond, MRS Proceedings, in press, (1997).

[33] K. Shiraishi, J. Phys. Soc. Jap. **59**, 3455 (1990).

[34] J. Neugebauer and M. Scheffler, Phys. Rev. B **46**, 16067 (1992).

[35] A. Wright and J. Nelson, Phys. Rev. B **50**, 2159 (1994); Phys. Rev. B **51**, 7866 (1995).

[36] K. Kim, W. Lambrecht and B. Segall, Phys. Rev. B **50**, 1502 (1994).

TABLE I. Calculated bulk properties of zinc-blende and wurtzite nitride semiconductors. The values of the gap at the Γ-pt (E_Γ) and of the valence-band width (ΔE_{vbw}) are the LDA results. Note that the LDA indirect gap in zinc-blende AlN is 3.2 eV. Experimental values are in brackets and follow Ref. 35.

	zinc-blende		
	AlN	GaN	InN
a_0 (Å)	4.37 (4.38)	4.52 (4.5)	5.01 (4.98)
B_0 (MBar)	2.02 (2.02)	1.70 (1.90)	1.58 (1.37)
	wurtzite		
a (Å)	3.09 (3.11)	3.20 (3.19)	3.55 (3.54)
c/a	1.62 (1.60)	1.63 (1.63)	1.63 (1.61)
u (units of c)	0.378 (0.382)	0.376 (0.377)	0.375
B_0 (MBar)	1.99 (2.02)	1.69 (1.95, 2.37)	1.62 (1.26, 1.39)

TABLE II. Elastic constants calculated for zinc-blende AlN, GaN and InN (in units of MBar). The experimental values in brackets are from Ref. 36.

	AlN	GaN	InN
c_{11}	3.01	2.61 (2.64)	2.14
c_{12}	1.62	1.27 (1.53)	1.37

FIG. 1. Schematic top view of the 2×2 gallium-adatom reconstruction on the (0001) surface of GaN. The gallium adatom (in grey) sits above a nitrogen in the second layer on the T_3 site. The numbers denote gallium atoms on the surface, while the primed numbers denote nitrogen atoms in the first sub-surface layer.

FIG. 3. Formation energy vs. gallium chemical potential for the (0001) surface. The maximum chemical potential for N (Ga) is equal to the energy per atom calculated for N_2 (bulk Ga). Two k-points are sampled in the irreducible Brillioun zone.
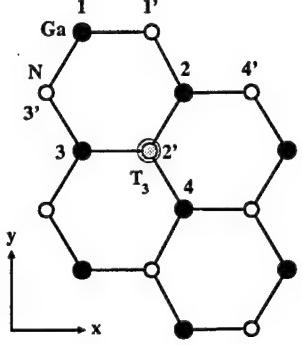
FIG. 2. Schematic top view of the 2×2 nitrogen-adatom reconstruction on the (0001) surface of GaN. The nitrogen adatom sits above the hollow (H_3) site. The numbers denote nitrogen atoms on the surface, while the primed numbers denote gallium atoms in the first sub-surface layer.

71
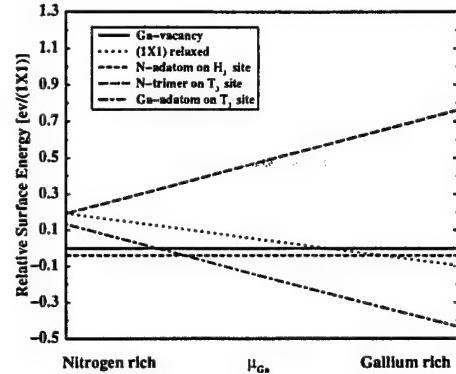
VI. Theory of Surface Morphology of Wurtzite GaN (0001) Surfaces

Krzysztof Rapcewicz, Marco Buongiorno Nardelli and J. Bernholc

Department of Physics, North Carolina State University, Raleigh, NC 27695-8202

Abstract

The influence of growth conditions and surface polarity upon the morphology of (0001) GaN surfaces is studied from first principles. The charge transfer between the Ga and N atoms in GaN and the very large electronegativity of nitrogen are found to play decisive roles in determining the stable reconstructions. Further, hydrogen stabilizes the ideally cleaved surface irrespective of polarity. For both polarities, adsorption of 3/4 of a monolayer of hydrogen results in a very stable surface with a 2×2 symmetry.

68.35.Md 71.55.Eq 31.15.Ar

The commercial production of high-brightness blue LED's and the recent demonstration of laser diodes,[1] have realized the expectations that have been held out for the III-V nitrides as optoelectronic materials. Advances in the application of this material to devices have been very closely tied to progress in the growth of epitaxial layers and heterostructures. Thus, a refined understanding of its growth is necessary for a further improvement of the quality of the material. In spite of the great effort that has been expended on the study of growth, even the polarity of the surface, *i.e.* whether the surface is Ga- or N-terminated, remains to be fully determined.[2] The surface morphology depends critically upon polarity,[3] and the uncertainty about the surface polarity will hinder progress in growth technology.

The two chief methods employed for epitaxial growth of GaN are metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE). Further, a variety of substrates and a spectrum of buffer layers have been used to grow GaN, making interpretation of the experimental data difficult. This is reflected in the fact that MBE grown samples of unknown polarity exhibit a 2×2 reconstruction,[4] while samples grown by MOCVD on sapphire have been shown to have different terminations depending upon the preparation of the buffer layer or growth parameters.[3,5,6] It is generally agreed that the Ga-terminated MOCVD-grown samples are smooth, which is consistent with the observed 2-dimensional growth in MBE of a presumed Ga-terminated surface. However, conflicting conclusions have been made about the N-terminated surface; it has been show to have inversion domains,[3] to be faceted[5] and to have a 1×1 reconstruction with a 3/4 monolayer of hydrogen coverage.[6]

In this situation, a theoretical characterization of the surface morphology can provide a framework for the interpretation of the experimental results. Thus, we have studied the dependence upon growth conditions of the 2×2 reconstructions of clean GaN(0001) surfaces with both polarities. We find that the ionic nature of GaN and the very large electronegativity of nitrogen determine the adsorption sites for the lowest energy adatom reconstructions and their subsequent behavior. Because the presence of hydrogen in MOCVD constitutes a major difference between the growth parameters of MBE and MOCVD, we have also studied the influence of adsorbed hydrogen on the surface morphology of GaN.

All calculations were carried out within the framework of a multigrid-based total-energy method that uses a real-space grid as the basis. Multigrid methods provide effective convergence acceleration and preconditioning on all length scales.[7] The Perdew-Zunger parametrization[8] of the Ceperley-Alder form[9] of the exchange-correlation energy was used. Non-local, norm-conserving pseudopotentials[10-12] were included using the Kleinman-Bylander approach.[13] The slabs contained four bilayers of GaN, three of which were relaxed. Relaxations in the third bilayer were negligible, indicating that a sufficiently large slab was employed. For calculations of the reconstructions of the Ga-terminated surface, pseudo-hydrogens of charge 0.75 passivated the N-terminated face, while for the study of the N-terminated face, pseudo-hydrogens of charge 1.25 passivated the Ga-terminated surface.[14] A vacuum region of 10 Å was used. An orthorhombic supercell with grid spacing of 0.27, 0.26 and 0.27 a.u. in the x-, y- and z-directions, respectively, was employed after a careful convergence study, thereby ensuring that the real-space description of the pseudopotentials is accurate.[15] Due to the well-known failure of the local density approximation to describe the band gap, we obtain a bulk LDA gap of 2.26 eV for GaN, as compared to the experimental value of 3.4 eV. However, the relative positions of the bulk and surface states are determined far more accurately. Further discussion of the calculations is given elsewhere.[16,17]

For the clean gallium- and nitrogen-terminated surfaces, we studied the 1×1 ideal, relaxed structure and seven different 2×2 reconstructions, namely the 2×2 vacancy structure, and 2×2 nitrogen adatom, gallium adatom and nitrogen-trimer structures, each on the T_3 and H_3 sites (see Figure 1). The lowest-energy reconstructions are all semiconducting and have surface states in the gap. We also investigated 1/4 and 3/4 monolayers of hydrogen adsorbed on the ideally cleaved surfaces of both polarities.

Fig. 2 displays the relative surface energies for the five lowest energy reconstructions of each type on the Ga-terminated (0001) surface. For gallium rich conditions, the gallium adatom sits on the T_3-site above the subsurface nitrogen atom. Because bonding in GaN has a strongly ionic component, charge is transferred from the Ga to the N atom. Consequently, the proximity of the gallium adatom to the nitrogen subsurface atom lowers the electrostatic

energy and makes this particular configuration more favorable energetically than the gallium adatom reconstruction with the adatom on the hollow site. Due to the stoichiometry of the gallium-adatom reconstruction on the (0001) surface, the dangling bond on the non-bonded gallium surface atom is empty. A surface state is nonetheless present at about 1.1 eV. This state corresponds to a bonding state between the gallium adatom and a gallium surface atom. The character of this bond is similar to the covalent bond seen in α-Ga.

Under nitrogen rich conditions, the nitrogen adatom sits on the hollow (H_3) site. This configuration is stabilized because it separates the nitrogen adatom from the subsurface nitrogen atom and thereby lowers the electrostatic energy. The presence of dangling bonds in the nitrogen-adatom reconstruction leads us to expect surface states in the gap. This is confirmed by the calculation, which shows surface states localized on the nitrogen adatom. There is a flat surface-band of p_x-p_y character that lies just below the valence-band maximum (VBM) throughout the whole Brillouin zone and decays into a resonance at Γ. We find that the true dangling-bond state of the nitrogen adatom lies about 1.0 eV above the VBM.

Fig. 3 shows the relative surface energies for the five lowest energy reconstructions of each type on the N-terminated (000$\bar{1}$) surface. Under gallium-rich conditions and for most of the range of the chemical potential, the reconstruction with the gallium-adatom on the H_3-site is the most energetically favorable. The adatom sits on the H_3-site. The same electrostatic argument applied to the Ga-terminated reconstructions is valid here: the gallium adatom on the hollow site optimizes the electrostatic energy by maximizing the separation between the gallium adatom and the gallium subsurface atoms. As a result of the stoichiometry of the gallium-adatom reconstruction, there is an occupied dangling-bond state on the non-bonded surface nitrogen atom. This gives rise to a surface state at about 0.3 eV above the VBM that has a p_z character.

Under nitrogen-rich conditions, a unique scenario occurs: the nitrogen adatom reconstructs to form an N_2 molecule which is bonded to a substrate that has a reconstruction very similar to the nitrogen vacancy reconstruction (see Figure 4), which, by itself, is unstable over the whole range of chemical potential (*cf.* Figure 3). The strength of the N-N bond

in the N_2 molecule is the driving force behind this reconstruction. Indeed, the dimerization of the nitrogen adatom is found to occur irrespective of our trial initial position of the extra nitrogen atom. This suggests that free nitrogen atoms in the vicinity of the nitrogen surface will prefer to bond to one of the surface atoms and form N_2 molecules. The binding energy of the N_2 molecule to the substrate is 0.3 eV, while the theoretical binding energy per nitrogen atom of N_2 is 5.9 eV. Consequently, the evaporation of N_2 from the surface is much more likely to occur than the dissociation of the molecule itself. As the vacancy reconstruction is unstable over the entire range of chemical potential, evaporation of the N_2 molecule will lead to a cascade of reconstructions. Correcting for the 1.1 eV/atom overestimate of the binding energy of the molecule by density functional theory will not alter this conclusion. A filled dangling-bond state with p_z symmetry is found just above the VBM in the case of the nitrogen-dimer/nitrogen-vacancy reconstruction. The molecular levels of the triple bond of the nitrogen dimer lie deep in the valence band, with a spectrum that resembles very closely the spectrum of the isolated molecule.

Characteristic of MOCVD growth is the presence of hydrogen. We have therefore studied how adsorbed hydrogen affects the surface morphology and find that its impact is significant. For both polarities, a surface with a 3/4 monolayer of hydrogen adsorbed on the ideally cleaved structure is more stable than any of the previously considered 2×2 reconstructions over the entire spectrum of the chemical potentials, where the ranges of the chemical potentials for gallium and nitrogen are bounded by the chemical potential for gallium metal and N_2 and the chemical potential of hydrogen is the binding energy per atom in the H_2 molecule.

For a 3/4 monolayer coverage on the Ga-terminated surface, the hydrogen atoms sit above the gallium atoms. As a result of the relaxation of the gallium atom not bonded to a hydrogen atom, the surface has a 2×2 symmetry. No surface states are found in the gap; however, there is a flat surface band at -0.9 eV, which decays into the bulk at Γ.

For a 3/4 monolayer of hydrogen adsorbed on the N-terminated surface, the hydrogen atoms sit above the nitrogen atoms. The H-N bond length is the very close to H-N bond

length in NH_3. The nitrogen atoms bonded to the hydrogen atoms undergo a small inward relaxation, while the nonbonded nitrogen relaxes very slightly outwards, so that the surface, in fact, has a 2×2 symmetry. As shown in Figure 5, this reconstruction has a surface band just below the valence band maximum. The band arises from the filled dangling-bond state on surface nitrogen atom that is not bonded to an adsorbed hydrogen atom and has p_z-symmetry.

LEED measurements on the N-terminated surface concluded that a surface with a 3/4 monolayer of hydrogen adsorbed has a 1×1 symmetry. The relaxation of the surface nitrogen atoms was found to be below the resolution of LEED,[6] which is in agreement with the small relaxations found here. Further a recent experiment on a sample grown by MBE on a sapphire substrate revealed a semiconducting surface with a 1×1 symmetry, which has a flat surface-band just below the valence band maximum.[18] We have compared these data to all of the stable surface structures described above. The only structure that reproduces the experimental results is the 3/4 monolayer H structure, in a 2×2 reconstruction on the N-terminated surface. In all other cases, the agreement is poor. In Fig. 4, the experimentally observed surface band is shown superimposed on the theoretically determined projected-band structure for 3/4 H-adatom structure on the N-terminated surface. The good agreement between our results and the data suggests that this sample is N-terminated with adsorbed hydrogen. However, it should be noted that the samples were grown by MBE and that hydrogen was not intentionally introduced into the system.[18]

In summary, we have investigated the dependence of the equilibrium surface morphology of GaN upon the growth conditions. The decisive factors favoring the lowest-energy reconstructions are the ionic character of the Ga-N bond and the strength of the N-N bond. Under gallium-rich conditions, gallium-adatom reconstructions are the most energetically favorable of the 2×2 reconstructions studied, irrespective of the polarity of the surface. Under nitrogen-rich conditions, a nitrogen-adatom reconstruction is most stable on the Ga-terminated surface. On the N-terminated surface, a nitrogen atom spontaneously bonds to a surface nitrogen atom creating a nitrogen molecule and a vacancy substrate. The molecule

is weakly bound and is expected to evaporate at the high temperatures of growth. Since the vacancy reconstruction is unstable over the whole range of chemical potential, the evaporation of the N_2 molecules should lead to a cascade of reconstructions of the substrate. For both polarities, a 3/4 monolayer of hydrogen stabilizes the relaxed, ideally cleaved surfaces; these reconstructions have a 2×2 symmetry and a flat surface band below the VBM. The very small relaxation of the surface atoms on the N-terminated surface, which is below the resolution of LEED,[6] means that this surface will appear as 1×1.

We would like to acknowledge fruitful conversations with Drs. P. Bogusławski, C. Bungaro and J. Schetzina.

REFERENCES

[1] S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku and Y. Sugimoto, Jpn. J. Appl. Phys. **35**, L74 (1996).

[2] G. Nowak, S. Krukowski, I. Grzegory, S. Porowski, J. M. Baranowski, K. Pakula and J. Zak, MRS Internet J. Nitride Semicond. Res. **1**, 5 (1996); E. S. Hellman, D. N. E. Buchanan, D. Wiesmann and I. Brener, MRS Internet J. Nitride Semicond. Res. **1**, 16 (1996); W. E. Packard, J. D. Dow, R. Nicolaides, K. Doverspike and R. Kaplan, Superlattices and Microstructure, **20**, 145 (1996).

[3] B. Daudin, J. L. Roviére and M. Arlery, Appl. Phys. Lett. **69**, 2480 (1996); J. L. Roviére, M. Arley, R. Niebuhr, K. H. Bachem and O. Briot, MRS Internet J. Nitride Semicond. Res. **1**, 33 (1996).

[4] W. C. Hughes, W. H. Rowland, Jr., M. A. L. Johnson, S. Fujita, J. W. Cook, Jr., J. F. Schetzina, J. Ren and J. A. Edmond, J. Vac. Sci. Technol. B **13**, 1571 (1995); and M. A. L. Johnson, Z. Yu, C. Boney, W. H. Rowland Jr., W. C. Hughes, J. W. Cook Jr., J. F. Schetzina, N. A. El-Masry, M. T. Leonard, H. S. Kong and J. A. Edmond, MRS Proceedings **449**, 271 (1997). T. D. Moustakas, Mat. Res. Soc. Symp. Proc. **395**, 111 (1996).

[5] F. A. Ponce, D. P. Bour, W. T. Young, M. Saunders and J. W. Steeds, Appl. Phys. Lett. **69**, 337 (1996).

[6] M. M. Sung, J. Ahn, V. Bykov, J. W. Rabalais, D. D. Koleske and A. E. Wickenden, Phys. Rev. B **54**, 14652 (1996).

[7] E. L. Briggs, D. J. Sullivan and J. Bernholc, Phys. Rev. B **52**, R5471 (1995); Phys. Rev. B **54**, 14362 (1996).

[8] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[9] D. Ceperley and B. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[10] G. Bachelet, D. Hamann and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[11] D. Hamann, M. Schlüter and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[12] D. Hamann, Phys. Rev. B **40**, 2980 (1989).

[13] L. Kleinman and D. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[14] K. Shiraishi, J. Phys. Soc. Jap. **59**, 3455 (1990).

[15] C. Bungaro, private communication, (1997).

[16] M. Buongiorno Nardelli, K. Rapcewicz and J. Bernholc, Phys. Rev. B **55**, R7323 (1997).

[17] K. Rapcewicz, M. Buongiorno Nardelli and J. Bernholc, to be published (1997).

[18] S. S. Dhesi, C. B. Stagarescu, K. E. Smith, R. Singh and T. D. Moustakas, preprint (1997).
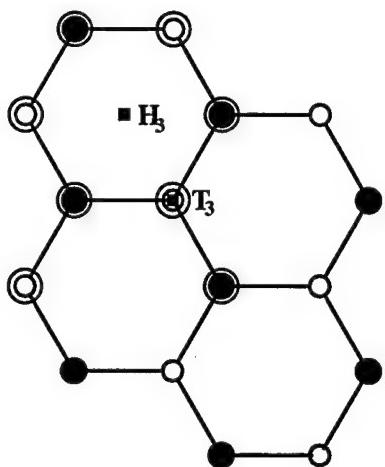
FIG. 1. Schematic top view of the (0001) GaN surface. On the Ga-terminated (0001) surface, the black atoms are gallium atoms; while the white atoms are nitrogen atoms; the situation is reversed on the N-terminated (000$\bar{1}$) surface. An atom in the T_3 site (represented by a square inside of a white circle) sits above a subsurface atom (a white circle), while an atom in the H_3 site (represented by a square) sits in the hollow site as shown. The circled atoms are in the 2×2 cell.
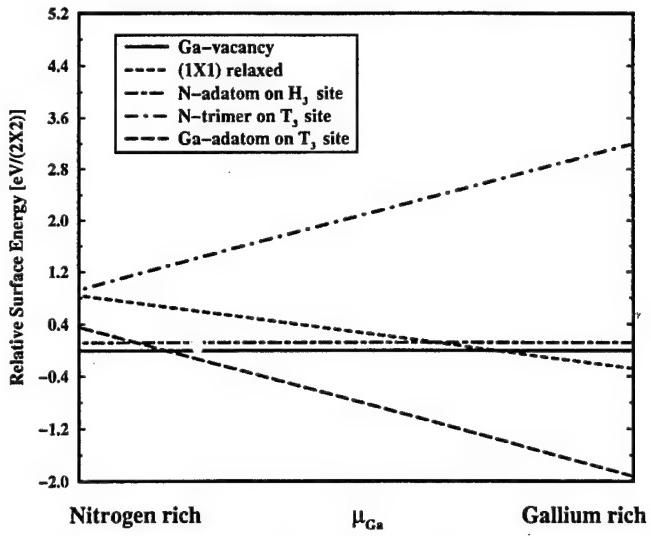
FIG. 2. Relative surface energy vs. gallium chemical potential for the (0001) Ga-terminated surface. The energies are calculated with respect to the 2×2 vacancy reconstruction. The maximum chemical potential for N (Ga) is equal to the energy per atom calculated for N_2 (bulk Ga).
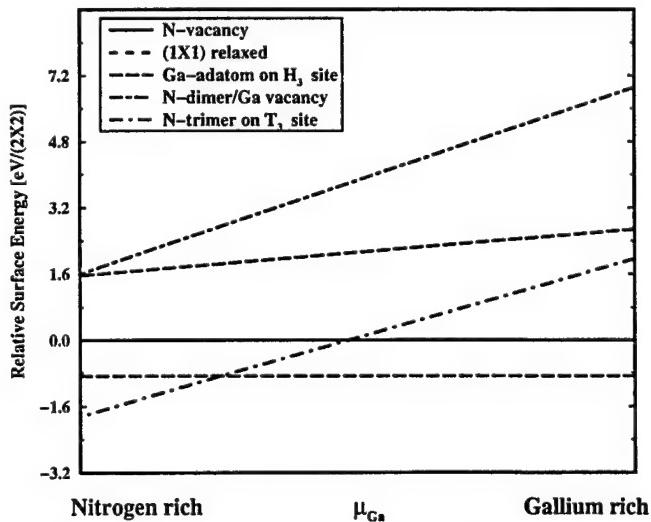
FIG. 3. Formation energy vs. gallium chemical potential for the nitrogen-terminated ($000\bar{1}$) surface. The energies are quoted with respect to the vacancy reconstruction. The maximum chemical potential for N (Ga) is equal to the energy per atom calculated for N_2 (bulk Ga).
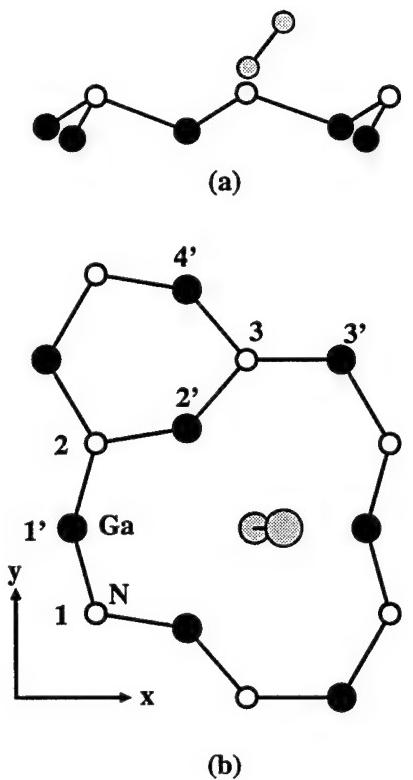
FIG. 4. Schematic side (a) and top (b) views of the 2×2 nitrogen dimer/nitrogen vacancy reconstruction on the (000$\bar{1}$) N-terminated surface of GaN. The reconstruction consists of a nitrogen dimer bound to a 2×2 nitrogen vacancy structure. The white circles denote the nitrogen atoms on the surface, while the dark circles denote gallium atoms in the first sub-surface layer. The shaded atoms comprise the N_2 dimer.
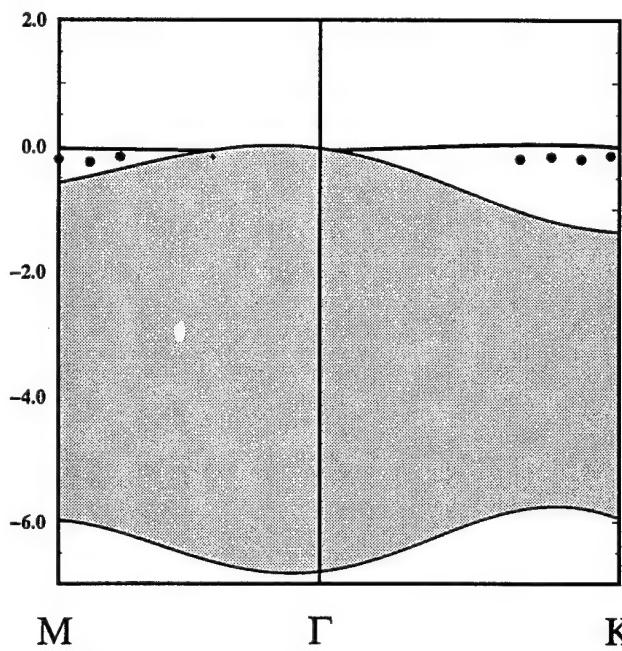
FIG. 5. Surface-projected band-structure for a 3/4 monolayer of hydrogen adsorbed on the nitrogen-terminated surface of GaN. The circles represent the experimental results of Ref. 18.

VII. Doping Properties of C, Si, and Ge Impurities in GaN and AlN

P. Bogusławski,[1,2] and J. Bernholc[1]

1. Department of Physics, North Carolina State University, Raleigh, NC 27695
2. Institute of Physics PAN, 02-668 Warsaw, Poland.

Doping properties of substitutional C, Si, and Ge impurities in wurtzite GaN and AlN were studied by quantum molecular dynamics. We considered incorporation of impurities on both cation and anion sublattices. When substituting for cations, Si and Ge are shallow donors in GaN, while Ge becomes a deep donor in AlN. Both impurities are deep acceptors on the N site. Substitutional C_{cation} is a shallow donor in GaN, but a deep one in AlN; C_N is a relatively shallow acceptor in both materials. Two effects that potentially quench doping efficiency were investigated. The first one is the transition of a donor from a substitutional position to a DX-like configuration. In crystals with the wurtzite symmetry, there are two possible variants of a DX-like state, and they have substantially different properties. In GaN, DX$^-$ states of both Si and Ge are unstable, or metastable, and thus they do not affect doping efficiency. In contrast, they are stable in AlN, and therefore neither Si nor Ge is a dopant in this material. Estimates obtained for $Al_xGa_{1-x}N$ alloys show that the crossover composition for DX stability is much lower for Ge ($x \simeq 0.3$) than for Si ($x \simeq 0.6$). The second effect quenching the doping efficiency is self-compensation, i.e., simultaneous incorporation of impurity atoms on both cation and anion sublattice. This effect may be enhanced by the formation of nearest-neighbor donor-acceptor pairs. The calculated binding energies of such pairs are large, about 1 eV, influencing self-compensation in some cases. Finally, the computed formation energies are used to identify growth conditions under which all these impurities may be efficient dopants in wide band gap nitrides.

I. INTRODUCTION

Wide band-gap nitrides are of considerable interest due to applications in blue/UV light-emitting diodes and lasers, and in high-temperature electronics [1,2]. To exploit fully the potential of these materials, understanding and control of doping needs to be achieved. In the present work we study substitutional X=C, Si, and Ge impurities in the hexagonal (wurtzite) GaN and AlN. This choice is motivated by the frequent usage of these species as dopants of III-V semiconductors [3–9]. Further, both C [10] and Si may be unintentionally incorporated as contaminants during growth.

In general, doping properties of group-IV atoms in a III-V compound are much more complex than those of group-II or group-VI atoms. This is because in the two latter cases the doping efficiency is determined by the electronic structure of dopant, and limited only by its solubility. However, a group-IV atom is likely to become a donor when incorporated on the cation site, and an acceptor on the anion site. Thus, a problem inherent to doping with group-IV elements is self-compensation, i.e., simultaneous incorporation of the dopant on both cation and anion sublattices. In GaN and AlN compounds, where there are large differences between the atomic radii of cations and anions, one could expect that self-compensation is blocked by strain effects. For example, a carbon atom in GaN should substitute for nitrogen, since the two atoms have similar atomic radii, while the substitution for the much bigger gallium induces a large lattice strain energy of a few eV. However, in competition with the strain effects are processes of electron transfer from donors to acceptors. Due to the wide band gap of nitrides, they lead to large energy gains thereby enhancing self-compensation. The tendency towards self-compensation is further increased by the formation of donor-acceptor pairs. According to our results, binding energies of nearest-neighbor donor-acceptor pairs are about 1 eV, due to both the Coulomb coupling and an additional short-range interaction of comparable strength.

Another factor which may affect the doping is a transition of X_{cation} donors from a substitutional to a DX-like state. In the wurtzite structure there are two nonequivalent DX-like configurations, in which the broken bond between the impurity and the host atom is either parallel or largely perpendicular to the c-axis. We have analyzed both variants and found that their properties are unexpectedly different in terms of stability and electronic structure. The transformation of the impurity to a DX configuration is commonly accompanied by a capture of a second electron by the donor, which quenches the doping efficiency. According to our results, this process does not occur for either Si or Ge in GaN (and indeed the quenching of the doping is not observed). In AlN, however, the DX$^-$ states are stable for both Si and Ge, implying that neither impurity is a dopant.

The present paper considers all of the above issues. This allows us to identify the conditions of growth under which C, Si, and Ge are excellent dopants, as well as the conditions for which substantial quenching of the doping efficiency should be expected. In addition, the studies of doping of pure GaN and AlN are extended to those of $Al_xGa_{1-x}N$ alloys. We find that these properties are strongly dependent on the alloy composition.

II. METHOD OF CALCULATIONS

The calculations were performed using quantum molecular dynamics [11], with atoms being efficiently relaxed using a special friction force [12]. Soft pseudopotentials for N and C were used [13], which allowed for a relatively low plane wave cutoff of 30 Ry. The pseudopotential of Ge was generated according to Ref. [14]. Impurities were placed in a 72-atom supercell. Due to size of the cell, the summations over the Brillouin Zone were approximated by one k-point (Γ). The d-electrons of Ga were treated as core states. This approximation leads to errors of about 0.2 eV in the band gap and 0.3 eV in the cohesive energy of zinc blende GaN [15]. In the calculations, the non-local core corrections were neglected. Our very recent results on native defects in $Al_x Ga_{1-x}N$ alloys, obtained using the potentials of Ref. [16], show that the formation energies computed with and without the non-local core corrections agree to within 0.2 eV. Further errors result from the use of the local density approximation, which is well-known to underestimate semiconductor band gaps, but these errors can be minimized by quoting the positions of the impurity-induced levels with respect to the nearest band edges. In particular, the effective-mass character found for several impurity-induced states is a feature that will persist in more accurate calculations.

III. SUBSTITUTIONAL CONFIGURATIONS AND STRAIN EFFECTS

We first discuss configurations of substitutional impurities and strain effects. A substitutional impurity in the crystal with the wurtzite symmetry has four nearest neighbors. One of them, located along the c-axis relative to the impurity (called here type-1 neighbor), is non-equivalent by symmetry to the remaining three neighbors (called here type-2 neighbors), see Fig. 1. For the group-IV impurities considered here, the non-equivalence effect is small, since the bond lengths with type-1 and type-2 neighbors differ by less than 1.5 % in most cases (the exceptions, C_{Al} and Ge_{Al}, are discussed below). In all cases we find breathing mode distortions that preserve the local hexagonal symmetry. The calculated changes of bond lengths relative to the bulk values (which are $b_1 = b_2 = 1.96$ Å for GaN, and $b_1 = 1.93$ and $b_2 = 1.89$ Å for AlN) are given in Table I. We also list energy gains E_{rel} due to the relaxation from the ideal substitutional configuration to the final one. As it is evident from the Table, the inclusion of relaxation effects is in general necessary for a proper description of impurities in nitrides. In cases when the mismatch between the atomic radii of impurity and the host atom is large, the calculated relaxation energies are one order of magnitude greater than those found

in typical III-V compounds like GaAs. The most drastic case is that of Ge_N in AlN; lattice relaxation releases an elastic energy of about 7 eV, increases the length of Ge-N bonds by 17 %, and rises the acceptor level by 1.3 eV. Very similar results are obtained for AlN:Si_N. The calculated E_{rel} are systematically greater for AlN than for GaN, reflecting the greater stiffness of AlN. Finally, we observe that even after the relaxation, *i.e.*, at equilibrium, the bonds around impurity remain strained (or stretched), which induces a residual strain energy. Simple arguments and estimates based on the Valence Force Field model [17] show that the residual strain energy is comparable to the relaxation energy E_{rel}. In other words, only one half of the initial strain is released during the relaxation. Large values of the residual strain suggest that formation of impurity-native defect complexes could be favored energetically. The implications of this effect for the stability of DX states are discussed in Sec. IV G.

IV. STRUCTURE OF DONORS

In this Section we consider the properties and electronic structure of C, Si, and Ge, substituting for the cation and becoming donors in both GaN and AlN. The discussion of results is complicated by the fact that, in many cases, the ground state configuration depends on the charge state of the impurity. We analyze two possible atomic configurations of a donor. The first one is substitutional, with almost equal bonds with its nearest neighbors. The second configuration is of DX-type [18]. In this case, a bond between the impurity and one of its first neighbors is broken, and one (or both) of these atoms move from substitutional sites to interstitial locations. Fig. 2 shows a schematic configuration diagram of an impurity, which takes into account both states, *i.e.*, the substitutional one and the DX-like state with a large lattice distortion. The case (a) corresponds to the situation when the DX state is unstable, and the impurity comes back to the substitutional state without an energy barrier. In the case (b), the DX state is higher in energy than the substitutional state, but there is an energy barrier between the two states and the DX configuration is metastable. The case (c) is similar to (b), but the energy of the DX state is lower than that of the substitutional configuration. Finally, in the case (d) the substitutional site is unstable, and the transformation to the DX state occurs with no energy barrier. As it will be discussed in the following, all the possible situations (a)-(d) are found for the group IV impurities in the nitrides.

An additional complication comes from the fact that in the wurtzite structure there are two non-equivalent DX configurations, for which the broken bond is between the impurity and either the type-1 or the type-2 neighbor. They are referred to as DX_1 and DX_2 variants in the following. We have considered both cases and found that

their properties differ substantially. In particular, the DX_1 variant is less stable than DX_2 for C^-, and more stable for Si^- and Ge^-. The properties of the impurities are qualitatively different in GaN and AlN, which is due in part to the wider band gap of the latter. Consequently, the impurity states in AlN are in general deeper and more localized, and the tendency to stabilize the DX-type geometry is more pronounced [19]. The results are summarized in Table II.

A. GaN:C

In the case of substitutional C in GaN, the C-derived level is a resonance situated at 0.8 eV above the bottom of the conduction band. Its position rises by 1.2 eV due to relaxation. For the neutral charge state of the impurity, the electron that should occupy the resonance autoionizes to the conduction band, and becomes trapped on the shallow level by the Coulomb tail of the impurity potential. However, the DX_2 and the substitutional configurations are energetically almost degenerate. We find that the energy of the DX_2 configuration is lower by about 0.1 eV than the substitutional one, but this small difference is within the accuracy of our calculations. The DX_1 variant, $C_{Ga}^{DX_1}$ is metastable, with energy higher than that of DX_2 by about 0.6 eV for all charge states. Therefore, carbon impurity in GaN corresponds to the case (b) and/or (c) in Fig. 2, since the two energy minima have almost the same energy.

The atomic positions for the DX-like configurations are shown in Fig. 1. In the DX_2 state, both the host N atom (denoted here by N^*) and the impurity are significantly displaced along the c-axis. The distance between both atoms is about 3 Å in the neutral charge state, which is almost twice the nearest neighbor distance (1.62 Å) in the substitutional geometry. The final positions of these atoms are close to the centers of 'triangles' formed by their neighbors. As in the DX_2 configuration, in the DX_1 variant both atoms are displaced from the lattice sites and approach the centers of triangles formed by their neighbors; the distance between the two atoms is again $\sim$ 3.2 Å. For both DX geometries, the $C-N^*$ distance is the largest for the negative charge states, and the shortest for positive ones. Considering the electronic structure, we find that C^{DX_1} (C^{DX_2}) introduces a singlet at about 0.3 (0.6) eV above the top of the valence band, and another singlet level at about 0.4 (0.5) eV below the bottom of the conduction band. Thus, C_{Ga} in GaN is a deep donor when in a DX configuration.

B. GaN:Si

Si_{Ga} in GaN is a shallow effective-mass donor. According to our results, Si is unstable in both DX variants,

since there are no energy barriers for the transition from the initial DX-like configurations, with large lattice distortions, to the substitutional ones. This holds for both the neutral and the negatively charged Si. Thus, Si_{Ga}^- is described by the case (a) in Fig. 2.

C. GaN:Ge

Ge_{Ga}, like Si_{Ga}, is a shallow effective-mass donor in GaN. Neither Si nor Ge introduce a resonance for energies up to 1.5 eV above the bottom of the conduction band. However, unlike Si, the negatively charged Ge is metastable in the DX_1 state, and its energy is higher by 0.6 eV than that of substitutional Ge_{Ga}^- (case (b) in Fig. 2.). For the neutral charge state, the DX_1 configuration is unstable, and the impurity comes back to the substitutional site with no energy barrier (case (a) in Fig. 2). We also find that the DX_2 variant is unstable for both the neutral and the negative charge states. The geometry of the DX_1 state is different than that obtained for C_{Ga}^{DX}. The Ge atom is essentially located at the lattice site, and only the nearest-neighbor N^* atom is displaced along the c-axis. The $Ge-N^*$ distance is 2.61 Å, to be compared with 1.93 Å for the substitutional configuration. $Ge_{Ga}^{DX_1}(-)$ introduces a singlet, almost degenerate with the valence band top, and a second singlet about 0.5 eV below the bottom of the conduction band.

D. AlN:C

The ground state configuration of carbon impurity in AlN depends on its charge state. For C_{Al} in the positive charge state, the ground state is the substitutional configuration, while the DX_1 and DX_2 variants are higher in energy by 0.15 and 0.3 eV, respectively. Similarly to GaN:C, for the neutral and the negative charge states the DX_2 variant is the ground state, and its energy is lower that that of DX_1 by 0.1 and 0.7 eV, respectively. Furthermore, for both the neutral and the negative charge states the substitutional configurations are unstable, and transform to the DX_1 states. This situation is schematically shown as case (d) in Fig. 2. The atomic configurations of C_{Al}^{DX} in both DX variants are similar to C_{Ga}^{DX} described above and shown in Fig. 1; both C and N^* move along the axis of the broken bond, and occupy positions close to the center of triangles of their respective neighbors. The distance between them increases to about 3.0 Å (depending on the charge state), which is almost two times larger than 1.62 Å found for the substitutional case.

$C_{Al}^{DX_1}(-)$ introduces a singlet at about 0.3 eV above the top of the valence band, and a singlet about 1.7 eV below the bottom of the conduction band. The levels introduced by $C_{Al}^{DX_2}$ are deeper, at about 0.9 eV above the valence band top, and at about 1.6 eV (for DX_2^+ and

DX_2^0) and 2.7 eV (for DX_2^-) below the bottom of the conduction band.

E. AlN:Si

The substitutional configuration is the ground state for both the positive and the neutral charge states of Si_{Al}. In this configuration, Si is an effective mass donor. The $DX_1(0)$ variant is higher by 0.3 eV than the ground state and the $DX_2(0)$ variant is unstable. However, for the negative charge state the ground state is the DX_1 variant, with the energy lower by 1.4 eV than that of the substitutional case. The $DX_2(-)$ state is metastable, and its energy is higher by 0.95 eV than that of $DX_1(-)$. The atomic configuration of the DX_1 variant for Si_{Al} in AlN is shown in Fig. 1. In this variant, the impurity remains on the substitutional site, the N* atom is strongly displaced above the triangle of its Ga neighbors, and the impurity-host distance increases to 3.0 Å, compared to 1.88 Å in the substitutional case. In the $DX_2(-)$ variant, both atoms are displaced from the ideal lattice sites, and the Si-N* distance is 2.93 Å. Si in the $DX_1(-)$ variant introduces a level at about 0.2 eV above the valence band top, and at about −1.9 eV below the bottom of the conduction band.

F. AlN:Ge

$Ge_{Al}(0)$ is a deep donor, with the energy level located at about 1 eV below the bottom of the conduction band. The localization of the wave function on Ge is accompanied by a large distortion of the atomic configuration around this donor; the nearest N neighbors relax outwards increasing the b_1 bond by 17%, and b_2 by 10% relative to the bulk values. This contrasts with the small relaxation found for Ge_{Ga} which is a shallow donor. For the neutral charge state, both DX variants are unstable. However, in the negative charge state Ge relaxes from the substitutional site towards the DX_1 configuration. The $DX_2(-)$ variant is unstable. The atomic configuration of the DX_1 variant for Ge_{Al} is similar to this for $Si_{Al}^{DX_1}$ in AlN shown in Fig. 1. In the DX_1 state the Ge atom occupies the substitutional site and the N* atom is displaced. Ge in the $DX_1(-)$ variant introduces a level at about 0.2 eV above the valence band top, and a very deep level at about −2.4 eV below the bottom of the conduction band. Finally, the calculated equilibrium configuration of $Ge_{Al}(0)$ closely resembles the D^- states in III-V zinc-blende semiconductors with large *breathing-mode* relaxations [20]. Considering the issue of D^-, we observe that a state with a large breathing mode distortion is a stable configuration of Ge in AlN, but in the neutral charge state. However, in the negative charge state this configuration is unstable, and the impurity transforms to

a DX state. In GaAs, the D^- state is more stable for Ge than for Si [20], we have thus not analyzed this state for Si.

From the results of Subsections A-F it follows that in all cases considered here, the breaking of an impurity-host bond leads to the formation of two levels in the band gap. To analyze structures of defect-induced levels we have projected their wave functions onto atomic orbitals [21]. The first level, denoted here by L_1, is energetically close to the top of the valence band. It is occupied by two electrons for all charge states of the impurity. It is localized on the displaced N* atom, although the degree of localization strongly depends on the system, displaying the ability of valence electrons to screen the broken bond. The second, higher level, denoted here by L_2, is pulled down from the conduction bands to the band gap. The L_2 level of the DX_1 variant is composed of comparable amounts of the s and p_z orbitals of the impurity, the sp^3 orbitals of the three N nearest neighbors of the impurity, and the p_z orbital of N*. In the DX_2 variant, the wave function of L_2 is composed from the s and p orbitals of the impurity, and of the triangle of N atoms around the impurity. In contrast to DX_1, the dangling bond of N* does not contribute to L_2.

The properties of DX states of Si and Ge are different from those of C. This difference may be traced back to the respective electronic structures of the impurities. In the case of Si or Ge, the degree of localization of L_1 is weak, since only about 10-15 % of the wave function is localized on N*. To complete the screening of broken bonds and stabilize the DX configurations, two electrons must occupy the L_2 level (which is much more localized than L_1). In the case of C, however, the L_1 level is highly localized. The contribution of N* orbitals is 35 % in the DX_1 variant, and two times greater in the more stable DX_2 variant. Thus, broken bonds are screened by the L_1 state to a large extent, and the occupation of L_2 is not necessary to stabilize the DX configurations.

Another factor adding to the stability of carbon in DX configurations relative to the substitutional ones comes from strain effects. When C substitutes for the much bigger Ga or Al atom, the C-N bonds are highly stretched, because they are shorter than the equilibrium cation-N bonds by about 15 %. This induces a high excess strain energy even after relaxation, since the C-N bonds remain stretched. The value of this excess elastic energy, estimated by the Valence Force Field model [17], is about 3 eV. (Note that this value is comparable to the cohesive energy per bond in the host crystals.) However, in the DX configurations one C-N bond is broken, and both C and N* are free to relax and release part of the excess strain by shortening the bonds with their neighbors. Consequently, the C_{cation}^{DX}-N* bonds are shorter by about 5% than the C_{cation}-N bonds in the substitutional case. Moreover, C^{DX} forms a nearly planar configuration with its neighbors, see Fig. 1. This geometry is additionally

stabilized by the tendency of C to form planar sp^2 bonds. We notice that both factors contribute to the stabilization of C_{cation}^{DX} in the neutral charge state, and even of the positive C_{Ga}.

G. Stability of DX states in $Al_xGa_{1-x}N$ alloys

The predicted stability of DX⁻ states for Si and Ge in AlN has important implications for the efficiency of n-type doping using these species. The stability of DX⁻ states implies that the reaction

$$2d^0 \rightarrow d^+ + DX^- \tag{1}$$

is exothermic. In this case, the electrons are captured on deep impurity levels (*i.e.*, on the L_2 states discussed above), and the doping does not result in a conducting sample. According to our results, this occurs for both Si and Ge in AlN. On the other hand, Si and Ge in GaN are excellent dopants since the DX states are metastable or unstable, and the reaction (1) is endothermic. These species are thus shallow donors. These results imply that the doping efficiency of both Si and Ge in $Al_xGa_{1-x}N$ alloys should strongly depend on the alloy composition. In GaN-rich samples Si and Ge are efficient donors, while in the AlN-rich limit the doping efficiency is quenched [22].

To determine the crossover compositions above which the low-energy DX_1 variant is stable, we have performed calculations for $Al_xGa_{1-x}N$ alloys with x=0.25, 0.5, and 0.75. The positions of the Ga and Al atoms in the supercell were chosen randomly. However, even in a random alloy, properties of a given impurity atom depend on the actual number and distribution of Ga and Al atoms in its surroundings, and in particular in the first two shells of its neighbors. This holds in particular for the relative stability of a DX configuration compared to the substitutional case. In the case of Si or Ge substituting for a cation, the first neighbors of the impurity are N atoms. Therefore, since the chemical disorder occurs only in the second-neighbor (and more distant) shells of the impurity, we do not expect a substitutional donor to be very sensitive to the actual surroundings. However, the DX_1 states may be sensitive to the chemical disorder. This is because the displaced atom is not the impurity but the host N* atom. In a DX state, N* is situated close to the center of a triangle of three cations, which are *first* neighbors of N*, and may be 3 Ga atoms, 2 Ga and 1 Al, *etc.*

We have analyzed this effect for both Si and Ge by considering a few configurations of cations for a given alloy composition. The calculated results are shown in Fig. 3, where we compare energies of DX_1 states with those of the corresponding substitutional cases [23]. We see that the relative energy of the DX state varies by about 0.4

eV depending on the actual environment of the impurity. In most cases, lowering the number of Ga atoms in the 'triangle' of first neighbors of N* stabilizes the DX state. The calculated magnitude of the alloy splitting of DX states is somewhat higher than that measured for $Al_xGa_{1-x}As$:Si [24]. Interestingly, the impurity-induced L_1 and L_2 levels are less sensitive to the chemical disorder than the total energy, since they remain constant to within ~0.05 eV for a given alloy composition.

As a result of the chemical disorder, the transition from the Ga-rich alloys where the DX states are metastable (or unstable), to the Al-rich alloys with the DX states stable will not occur at a sharp crossover composition. Instead, one expects a finite composition range of the order of 10%, over which the doping efficiency will gradually decrease. From Fig. 3 it follows that the crossover compositions of $Al_xGa_{1-x}N$ are about 0.3 for Ge, and about 0.6 for Si. Thus, from this point of view, Si is a donor superior to Ge, since it remains an efficient dopant over a wider composition range.

H. Comparison with experiment

We will now compare our results for C, Si, and Ge impurities with the available experimental data. Recent investigations have shown that C is an acceptor slightly more shallow, by about 20 meV, than the commonly used Mg [4]. To date, a hole concentration of $\sim 3x10^{17}$ cm⁻³ was obtained using doping with CCl_4 in MOMBE growth [6]. Higher concentrations were not obtained since with the increasing CCl_4 flow the growth rate is drastically reduced. One should note, however, that the theoretical solubility of C_N is one order of magnitude higher.

Experimental investigations of Si in GaN were recently performed by Rowland *et al.* [7]. They have shown that up to the concentrations of $4x10^{19}$ cm⁻³, Si is incorporated without self-compensation. The high-pressure study of Wetzel *et al.* [8] shows that Si remains a shallow donor for pressures up to 27 GPa. This suggests an absence of a resonant state close to the bottom of the conduction band, and indicates that Si is a hydrogenic donor in $Al_xGa_{1-x}N$ alloys with the composition range $0 \leq x \leq 0.3$. No conclusions for higher values of x can be drawn. This also indicates that the possible DX state is much higher in energy than in the case of GaAs, where the DX state of Si is stable at pressures above 2 GPa [24]. Bremser and Davis [5] have recently observed quenching of doping efficiency with Si for compositions higher than 60 per cent, which may be due to the stabilization of DX states. Both experimental findings [8,5] agree with our calculated properties of Si impurity.

Experimental investigations of the $Al_xGa_{1-x}N$:Ge were recently performed by Zhang *et al.* [9]. They find that Ge is an efficient dopant for x≤0.2, which covers

the whole composition range of efficient doping with Ge predicted here. No data for higher Al content were given.

We end this Section with two comments. First, we observe that well-known fingerprints of the (meta)stability of DX states are the so-called persistent effects, *e.g.*, persistent changes in conductivity induced by illumination of a sample cooled in darkness. They occur when both the substitutional and the DX configurations of a negatively charged impurity are local minima separated by an energy barrier, which allows for a metastable, non-equilibrium distribution of electrons between the two states to exist. This corresponds to the cases (b) and (c) in Fig. 2. One should observe that the conditions for the persistent effects can be different from those leading to quenching of the doping efficiency, which takes place in situations (c) and (d) in Fig. 2. Thus, the two effects take place in different regimes of alloy composition. According to our results, two minima exist for C, Si, and Ge in $Al_xGa_{1-x}N$ alloys with appropriate compositions, which may be estimated from Table II and Fig. 3. In particular, persistent effects for Ge and C may only occur in the Ga-rich regime, since in AlN the substitutional configurations are unstable, and both impurities undergo a transformation to the DX_1 variant without an energy barrier. Si^- in GaN and Ga-rich $Al_xGa_{1-x}N$ alloys is unstable in the DX state, and thus persistent effects may only occur for Al contents higher than about 25 %. To our knowledge, such effects were not reported for $Al_xGa_{1-x}N$ alloys for the dopants considered here.

The second comment concerns shallow effective mass donors in $Al_xGa_{1-x}N$ alloys. The doping efficiency of substitutional Si and Ge decreases with an increasing Al content because, as was observed experimentally, the effective-mass states become progressively deeper with the increasing alloy composition due to both the increasing effective mass and the decreasing dielectric constant. We stress that this effect is independent of the stabilization of the DX states with the increasing Al content.

V. ELECTRONIC STRUCTURE OF ACCEPTORS

We shall now consider the case when group-IV atoms substitute for N and form acceptors. The electronic structure of acceptors is more complex than that of donors. In zinc-blende materials, acceptor levels of group-IV atoms are three-fold degenerate due to the cubic symmetry of the host. In GaN and AlN with wurtzite symmetry, the crystal field splits the triplet into a doublet and a singlet. Here, the energy of doublet-singlet splitting is denoted by E_{split}. In all cases considered here, the energies of doublets, E_{DB}, are higher than those of singlets, E_S. For the neutral charge state of acceptors, the singlet is occupied by two electrons, and the doublet by three electrons. Both the energy levels and the splitting energies strongly depend on the impurity. The calculated

values of E_{DB} for neutral acceptors are listed in Table I. We find that C is a shallow acceptor with E_{DB}=0.2 eV in GaN, and deeper (E_{DB}=0.5 eV) and more localized in AlN. The doublet-singlet splitting E_{split} is about 0.2 eV in both materials. In contrast, both Si and Ge are deep acceptors. For Si(Ge) in GaN we find E_{DB}=1.2 (1.35) eV and E_{split}=0.6 (0.6) eV. In AlN the binding energies are even higher, with E_{DB}=2.0 eV and E_{split}=0.65 eV for Si_N, and E_{DB}=2.1 eV and E_{split}=0.65 eV for Ge_N. From these results it follows that the incorporation of Si and Ge on the N sites does not result in p-type samples; this situation is dissimilar to that in, *e.g.*, GaAs, where both Si and Ge are shallow acceptors.

According to our results, the energy levels of Si and Ge are similar, but they are qualitatively different from those of C. These differences are due to both different atomic energies of p orbitals, which are about 1 eV lower for C than for Si and Ge, and differences in atomic radii, which affect both atomic relaxations and the energies of the impurity states.

VI. FORMATION ENERGIES AND COMPENSATION EFFECTS

The concentration of an impurity at thermodynamic equilibrium is given by

$$[conc] = N_{sites}exp(S_{form}/k_B - E_{form}/k_BT), \quad (2)$$

where N_{sites} (equal to 36 atoms in a 72-atom cell) is the concentration of atomic sites and S_{form} is its formation entropy, which is typically about 4-8 k_B.

The formation energy of, *e.g.*, Si in GaN in a charge state q is

$$E_{form}(q) = E_{tot}(q) - n_{Ga}\mu_{Ga} - n_N\mu_N - \mu_{Si} + qE_F,$$
$$(3)$$

where E_{tot} is the total energy of the supercell with the impurity, n_{Ga} and n_N are the numbers of Ga and N atoms in the supercell, μ is the chemical potential, and E_F is the Fermi energy. The chemical potentials depend on the source of atoms involved in the process, and therefore on the actual experimental situation. At equilibrium,

$$\mu_{Ga} + \mu_N = \mu_{GaN} = \mu_{Ga(bulk)} + \mu_{N_2} + \Delta H_f(GaN), \quad (4)$$

where ΔH_f is the heat of formation, which is negative for a stable compound. The highest possible concentration of a given dopant is obtained at the site of its lowest formation energy, see Eq. (2), and at the highest level of the chemical potential at the source of the dopant atoms. For C doping, we have assumed that the source is an elemental solid, *i.e.* graphite, which sets the correct upper bound for μ_C. Consequently, the values of the formation energies of both C_{Ga} and C_N change by the value of

the heat of formation of GaN between the two extremal conditions of growth, the Ga-rich and the N-rich limits. For Si and Ge doping, the possible formation of Si_3N_4 and Ge_3N_4 stable alloys lowers the upper bound of μ due to an additional constraint [25] on the chemical potentials, *e.g.*:

$$3\mu_{Si} + 4\mu_N = \mu_{Si_3N_4} = 3\mu_{Si(bulk)} + 4\mu_{N_2} + \Delta H_f(Si_3N_4).$$
$$(5)$$

By combining Eqs. (3-5) we find that the chemical potential of Si depends on the conditions of growth:

$$\mu_{Si} = \mu_{Si(bulk)} + (1/3)\Delta H_f(Si_3N_4) - (4/3)\Delta H_f(GaN)$$
$$(6)$$

for Ga-rich limit, and

$$\mu_{Si} = \mu_{Si(bulk)} + (1/3)\Delta H_f(Si_3N_4) \qquad (7)$$

for the N-rich limit. Therefore, for the N-rich case, the upper bound of μ_{Si} is always reduced. This result is due to the fact that in the presence of excess nitrogen silicon atoms form the stable compound Si_3N_4. The upper bound of μ_{Si} is reduced form its bulk value also for the Ga-rich case (Eq. 7), since

$$(1/3)\Delta H_f(Si_3N_4) - (4/3)\Delta H_f(GaN) < 0. \qquad (8)$$

However, for GaN:Ge, AlN:Si, and AlN:Ge in the cation-rich limit, the relation analogous to (8) is not fulfilled, and the corresponding upper bounds are not reduced from their bulk values.

The calculated formation energies are given in Table III [26]. As follows from the Table, C is preferentially incorporated on the anion sublattice (except for the N-rich limit in AlN, where the formation energies of C_{Al} and C_N are comparable), while both Si and Ge substitute for cations. These results are simply explained by the presence of residual strain energy when the mismatch between the host and the impurity atoms is large. In these cases, the excess strain energy is of the order of a few eV, as pointed out in Sec. III. However, as we show below, charge transfer effects may overcome this effect.

Formation energies listed in Table III are those of neutral impurities. In the case of charged impurities, formation energies are in general reduced by electron transfer to/from the Fermi level according to Eq. (3). Since the energy gain may be of the order of the band gap, the formation of charged defects is of particular importance in wide band gap materials. (This situation is similar to that of native defects, discussed recently in Ref. [27].) For the dopants considered here, the electron transfer effects may overcame the strain-driven preference of the dopant atom X to substitute for the component of the similar size, rendering formation energies of X_{cation}

and X_N close to each other. Consequently, a simultaneous incorporation of X on both sublattices (*i.e.*, self-compensation) becomes possible. However, as we show below, the degree of self-compensation strongly depends on the conditions of growth, and may vary from none to total self-compensation.

Defect concentrations are also very sensitive to the growth conditions, i.e., the temperature and the chemical potentials of the species involved. For this reason, and for the sake of transparency, we have performed calculations for a limited set of parameters, choosing various extremal conditions of growth in order to assess when an efficient and non self-compensated doping is expected at all.

From Table III, for Si and Ge in the N-rich conditions of growth the difference of formation energies between X_N and X_{cation} exceeds the value of the band gap. This holds for both GaN and AlN. Therefore, under these conditions, energy gain due to charge transfer from donor to acceptor cannot overcome strain-driven effects, and the concentration of X_{cation} is always higher than that of X_N, *i.e.*, self-compensation is negligible. We observe, however, that in this limit X_{cation} donors can still be compensated by the dominant native acceptors, which are cation vacancies [27]. In particular, the calculated solubility limit of Si at 800 °C exceeds 10^{20} cm^{-3}, but the degree of compensation by V_{Ga} is high, because formation of cation vacancies is an efficient process in the N-rich conditions. Analogous arguments hold for doping of GaN with carbon. From Table III it follows that under cation-rich conditions of growth, the formation energies of C_{cation} are larger than these of C_N by more than the band gap. Thus, in this limit, carbon is a non-self-compensated acceptor. However, C_N may be compensated by dominant native donors, which are nitrogen vacancies and/or Ga interstitials [27,28].

On the other hand, under the opposite conditions of growth (*i.e.*, the cation-rich limit for Si and Ge, and the N-rich limit for C), self-compensation may play a dominant role. In the case of GaN:C, GaN:Ge, AlN:C, AlN:Si, and AlN:Ge, a value of the Fermi level exists for which $E_{form}(X_{cation}^+) = E_{form}(X_N^-)$, at which point the concentrations of X_{cation}^+ and X_N^- should be same. This implies that under equilibrium conditions doping with group IV atoms should lead to self-compensation and pinning of the Fermi level. For Si in GaN, however, self-compensation effects are not expected except in the very Ga-rich limit. In practice, kinetic effects may limit dopant incorporation and thus allow for reasonably effective doping.

We end this Section by comparing our results with other theoretical calculations for dopants in GaN. In Ref. [29], somewhat higher formation energies for both C_{Ga}, 6.5 eV, and Si_{Ga}, 1.2 eV, were obtained, but they are still in a satisfactory agreement with the present work [26]. Fiorentini *et. al.* [30] have studied potential accep-

tors in GaN. Considering C_N, they find that its acceptor level, E_{acc}=0.65 eV, is deeper than what is obtained here, and the formation energy E_{form}=4.2 eV is higher than our value of 2.8 eV. Also, Si_N with E_{acc}=1.97 eV, is predicted to be deeper by about 0.8 eV than what we find, while E_{form}=5.5 eV at N-rich conditions is lower than our value of 6.9 eV. A comparison of Table 1 from Ref. [30] with Table I of this work shows that the relaxation effects predicted here are somewhat larger for both dopants; e.g., for Si_N, Fiorentini et. al. find Δb=12.0 %, and E_{rel}=3.4, while our respective values are 13.6 % and 3.9 eV. These differences may be due to a larger unit cell used in the present work, which allows for a greater magnitude of atomic relaxations, and possibly differences in the atomic pseudopotentials. Si_{cation} donors have recently been investigated by Mattila and Nieminen [31]. They consider cubic GaN and AlN, which makes the comparison less direct. However, their results for GaN:Si are close to ours: Si_{Ga} is a shallow donor, and its E_{form} =1.6 eV is in an acceptable agreement with our value. The discrepancies are somewhat larger for Si in AlN. In particular, their E_{form} is lower than our value by about 1.0 eV. Moreover, they find Si to be a deep donor. The inward relaxation of nearest neighbors for Si^+ is similar to that found here. The pronounced asymmetry of the atomic relaxation around Si_{Al}^-, i.e., the elongation of the bond with one NN by 12 %, and the shortening of the three remaining bonds, may be a precursor of the stabilization of the DX configuration that we find in the wurtzite AlN; however, DX-like states with large lattice relaxations were not analyzed in Ref. [31].

VII. FORMATION OF NEAREST-NEIGHBOR DONOR-ACCEPTOR PAIRS

The degree of self-compensation may be further increased by the formation of donor-acceptor nearest-neighbor (NN) pairs. Formation of such pairs is an independent mode of incorporation of impurity atoms. In a simple picture which neglects the coupling between second-neighbor or more distant pairs, the total concentration of an impurity X is the sum of $[X_{cation}]$, $[X_N]$, and $[XX]$. The formation of NN pairs increases both the solubility (as was discussed for Si in GaAs in Ref. [32]) and the degree of self-compensation, but does not change the concentrations of conduction electrons or holes. However, the presence of the pairs may significantly lower carrier mobility since they introduce an additional scattering channel.

As the result of electron transfer from donors to acceptors in the pairs, both X_{cation} donors and X_N acceptors are ionized. According to our results, the binding energy of a second-neighbor, or a more distant pair is essentially the Coulomb interaction between two point charges embedded in the medium with the appropriate dielectric constant. However, the binding in a NN pair deviates significantly from the purely Coulombic character. This is not surprising, since the picture of interacting point charges should not hold for NN pair, given the finite extent of the impurity wave functions. We define the binding energy of a NN pair as

$$E_{pair} = E_{NN} - E_{\infty}, \qquad (9)$$

where E_{NN} and E_{∞} are the total energies of a NN pair and of a distant pair, respectively. The additional short-range interaction is defined as

$$E_{sr} = E_{pair} - E_{Coul}^{NN}, \qquad (10)$$

where E_{Coul}^{NN} is the Coulomb energy of a pair of point charges separated by the theoretical NN distance.

The calculated values of E_{pair} and E_{sr}, corrected for the spurious interaction between pairs in different supercells, are given in Table IV. We have only investigated the NN pairs oriented along the c-axis. The pairing energies are about 1 eV, which is higher than the Si-Si binding energy in GaAs [32,33], due in part to a smaller NN distance and a smaller dielectric constant. The additional short-range interaction E_{sr} is always attractive. It ranges from -0.15 eV for a Si-Si pair in GaN to -1.0 eV for a Ge-Ge pair in AlN (which is about 60 % of the binding energy). We also see that E_{sr} is systematically larger in AlN than in GaN by 0.3–0.6 eV. The attractive character of the short-range coupling is in part due to the release of excess strain, which is more efficient for NN than for more distant pairs.

The formation energy of a NN pair is

$$E_{form}(XX) = E_{form}(X_{cation}^+) + E_{form}(X_N^-) + E_{pair},$$
$$(11)$$

which is independent of the Fermi level E_F for neutral pairs, i.e., for $E_A < E_F < E_D$. The concentration of NN pairs is given by Eq. 2, with N_{sites} four times larger than that for substitutional impurities. Thus, $[XX]$ is comparable to $[X_{cation}]$ or $[X_N]$ only when $|E_{pair}| \geq max[E_{form}(X_{cation}^+), E_{form}(X_N^-)]$.

The concentration of NN pairs may be calculated based on the results from Tables I, III and IV. Due to the fact that C does not form a stable compound with N, the concentration of C-C pairs is independent of the conditions of growth. We find that pairing of C is negligible in GaN, but can be important in AlN, especially under N-rich conditions. For both Si and Ge in both GaN and AlN under the N-rich conditions, we find that the concentration of NN pairs is negligibly small compared with the concentrations of the isolated impurities. However, the pairing may be important in the cation-rich limit. In particular, for Si in GaN that the concentration of NN pairs is comparable to that of $[Si_{Ga}]$ and $[Si_N]$ in the Ga-rich limit. In AlN:Si, most of Si impurities form pairs

under the Al-rich conditions. Finally, in the case of Ge in GaN, [Ge-Ge] is smaller by one order of magnitude that [Ge$_{Ga}$] and [Ge$_N$], but in AlN the concentration of Ge pairs is close to that of isolated Ge impurities in the Al-rich limit.

The electronic structure of the nearest-neighbor X$^+_{cation}$-X^-_N pairs is similar to that of distant X$_{cation+}$ and X^-_N impurities, the main modification being a factor of two increase in the doublet-singlet splittings. More specifically, in GaN E$_{split}$ is about 0.4, 1.0, and 1.05 eV for C-C, Si-Si, and Ge-Ge pairs, respectively. For AlN, the splitting increases to about 0.7, 1.4, and 1.35 eV for C-C, Si-Si, and Ge-Ge pairs, respectively. This effect is due to the close proximity of the X$_{cation}$ donor, located along the c-axis relative to the acceptor X$_N$. The presence of the donor enhances the non-equivalence between the z-symmetry singlet, pulled down by the donor potential, and the (x, y)-symmetry doublet.

VIII. SUMMARY

In summary, we have investigated doping properties of C, Si, and Ge in wurtzite GaN, AlN, and Al$_x$Ga$_{1-x}$N random by Quantum Molecular Dynamics. Incorporation of impurities on both cation and N sublattices was considered. We have analyzed effects that could potentially limit the doping efficiency, namely: (i) stabilization of donors in DX configurations, (ii) self-compensation due to the simultaneous incorporation of impurities on both sublattices, and (iii) the formation of nearest-neighbor donor-acceptor pairs. In cases when the atomic radii of the impurity and of the substituted host atom are substantially different (e.g. C$_{Ga}$, or Si$_N$), the effects of atomic relaxations around the impurities and the induced strain strongly affect both the impurity levels and the formation energies. However, the strain-driven preference to substitute for the host atom with a similar size may be overcome by the energy gain due to electron transfer between a donor and an acceptor. If this is the case, doping at equilibrium conditions should lead to Fermi level pinning, at which point the formation energy of the donor becomes equal to that of the acceptor, and thus total self-compensation takes place.

We find that

(i) C$_N$ is a promising acceptor, as it is somewhat more shallow than the commonly used Mg [34,4] and it should have a significantly greater solubility than atomic Mg. However, the presence of H aids in incorporation of Mg [35,36], although in this case post-growth activation of Mg is required.

(ii) Negligible self-compensation and pairing effects are expected for Al$_x$Ga$_{1-x}$N alloys in the Ga-rich limit, i.e., for compositions $0 < x < 0.4$.

(iii) in Al$_x$Ga$_{1-x}$N alloys with composition $x < 0.60$ grown in the N-rich conditions, Si$_{cation}$ is an excellent effective-mass donor; neither self-compensation nor the formation of NN pairs is expected. For alloys with compositions higher than about 0.60, the DX$^-_1$ configuration is stable, which quenches the doping efficiency.

(iv) for the N-rich conditions of growth, Ge$_{cation}$ is an excellent effective-mass donor in Al$_x$Ga$_{1-x}$N for compositions lower than about 0.3. For $x > 0.3$, the doping efficiency is quenched due to the stabilization of the DX$^-_1$ state. Furthermore, appreciable self-compensation and pairing are expected.

Samples grown under the conditions opposite to those specified above are expected to be self-compensated to a high degree. Our results for Si and Ge indicate that the two common n-type dopants in GaN, Si and Ge, are not shallow donors in AlN and Al-rich alloys; this opens the question of finding not only efficient acceptors, but also appropriate donors for Al-rich materials.

Note added in Proof:
A paper [37] that appeared after the present work was submitted, also investigated the stability of DX states of Si and Ge dopants in wurtzite GaN and AlN. This paper analyzed configurations with the broken bond parallel to the c-axis (calling it γ-BB, an analog of our DX$_1$ state), and along one of the three remaining bonds (called α-BB, an analog of our DX$_2$ state). The γ-BB configuration considered in Ref. [37] had a displaced Si or Ge impurity atom and an almost non-displaced host N atom. It was found unstable for both Si and Ge in Al$_x$Ga$_{1-x}$N alloys. This is in in agreement with our results. The configurations DX$_1$ that we have considered have the reverse displacements, with the impurity atom almost non-displaced and the host N* strongly displaced along the c-axis (see Fig. 1 and Sec. IV). This variant turned out to be metastable (for Ge in GaN), or even stable (for both Si and Ge in AlN, see Table II).

Turning to the α-BB (DX$_2$) variant, both we and Ref. [37] find it unstable for Ge in Al$_x$Ga$_{1-x}$N for all compositions. However, for Si the agreement is less satisfactory: Ref. [37] finds that this variant is metastable in GaN, and stable in AlN. We find that this variant is much less stable: in GaN it is unstable, and in AlN it is less stable than DX$_1$ by 0.95 eV. Our results imply that the shallow-deep transition occurs at a higher Al concentration (cf. Section 4H), which appears to be supported by recent experiments [5]. Finally, an impurity may have several local minima with large lattice relaxations, as was found for CdTe:Cl [38], but the configuration shown in Fig. 3d of Ref. [37] is too sketchy to allow for a detailed comparison with our results.

[1] R. F. Davis, Physica B **185**, 1 (1993).

[2] H. Morkoc *et al.*, J. Appl. Phys. **76**, 1363 (1994).

[3] S. Nakamura, T. Mukai, and M. Senoh, Jpn. J. Appl. Phys. **31**, 2883 (1992).

[4] S. Fisher, C. Wetzel, E. E. Haller, and B. K. Meyer, Appl. Phys. Lett. **67**, 1298 (1995).

[5] M. Bremser and R. F. Davis, unpublished.

[6] C. R. Abernathy, J. D. MacKenzie, S. J. Parton, and W. S. Hobson, Appl. Phys. Lett. **66**, 1969 (1995).

[7] L. B. Rowland, K. Doverspike, D. K. Gaskill, Appl. Phys. Lett. **66**, 1495 (1996).

[8] C. Wetzel, A.L. Chen, T. Suski, J.W. Ager II, and W. Walukiewicz, Physica Status Solidi, to appear.

[9] X. Zhang, P. Kung, A. Saxler, D. Walker, T.C. Wang, and M. Razeghi, Appl. Phys. Lett. **67**, 1745 (1995).

[10] G.-C. Yi and B. Wessels, Appl. Phys. Lett. **70**, 357 (1997).

[11] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[12] C. Wang, Q.-M. Zhang, and J. Bernholc, Phys. Rev. Lett. **69**, 3789 (1992).

[13] G. Li and S. Rabii, unpublished (1992).

[14] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[15] V. Fiorentini, M. Methfessel, and M. Scheffler, Phys. Rev. B **47**, 13353 (1993).

[16] M. Buongiorno Nardelli, K. Rapcewicz, and J. Bernholc, Phys. Rev. B **55**, R7323 (1997).

[17] P.N. Keating, Phys. Rev. **145**, 637 (1966); R.M. Martin, Phys. Rev. B **1**, 4005 (1970). The bond-stretching and the bond-bending constants were obtained by scaling the values used by A. B. Chen and A. Sher, Phys. Rev. B **32**, 3695 (1985) for other III-V semiconductors. For GaN(AlN) we use $\alpha = 80$ (89) N/m, and $\beta = 12$ (13) N/m.

[18] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).

[19] The effects of the band structure and chemical bonding on the stability of DX states in GaAs and InP were discussed by B.-H. Cheong and K.J. Chang, Phys. Rev. Lett. **71**, 4354 (1993).

[20] T. M. Schmidt, A. Fazzio, and M. J. Caldas, Phys. Rev. B **53**, 1315 (1996).

[21] P. Bogusławski and I. Gorczyca, Semicond. Sci. Technol. **9**, 2169 (1994).

[22] Strictly speaking, even in the case (c) of Fig. 2 a sample may be conducting at finite temperatures. The progressive decrease of the energy of the DX state with the alloy composition investigated in this Section results in an increasing activation energy of the conduction electrons. In addition, the conductivity may be due to persistent effects discussed in the following.

[23] We note that even in the absence of the impurity, total energies of supercells with different distributions of cations are in general different. Thus, to obtain energy of DX state relative to the substitutional one, one has to analyze both states for each distribution of cations.

[24] Z. Wilamowski, J. Kossut, W. Jantsch, and G. Ostermayer, Semicond. Sci. Technol. **6**, B38 (1991).

[25] C. G. Van De Walle, D.B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **47**, 9425 (1993).

[26] The results of Table III were obtained using theoretical values of ΔH_f of the considered materials, except for Si_3N_4 and Ge_3N_4. Cohesive energies of these two compounds were not computed. Instead, we used values based on experiment; to correct for the usual overestimation of formation energies in local density theory, we increased the experimental values by 20%, which is the average discrepancy between the computed and experimental values for bulk Si, Ge, and N_2. This correction of ΔH_f results in a 0.5 eV increase of E_{form} for Si in both GaN and AlN in the N-rich limit (Eq. 7). The corresponding increase in the formation energy for Ge is 0.3 eV.

[27] P. Bogusławski, E. L. Briggs, and J. Bernholc, Phys. Rev. B **51**, 17255 (1995).

[28] P. Perlin, T. Suski, H. Teisseyre, M. Leszczynski, I. Grzegory, J. Jun, S. Porowski, P. Bogusławski, J. Bernholc, J. C. Chervin, A. Polian, and T. D. Moustakas, Phys. Rev. Lett. **75**, 296 (1995).

[29] J. Neugebauer and C. Van de Walle, in Proc. 22 Int. Conf. Phys. Semicond. Vancouver 1994, p. 2327, ed. D. J. Lockwood (World Scientific 1995).

[30] V. Fiorentini, F. Bernardini, A. Bosin, and D. Vanderbilt, in Proc. 23 Int. Conf. Phys. Semicond. Berlin 1996, p. 2877, ed. M. Scheffler and R. Zimmerman (World Scientific 1996).

[31] T. Mattila and R. M. Nieminen, Phys. Rev. B **55**, 9571 (1997).

[32] J. E. Northrup and S. B. Zhang, Phys. Rev. B **47**, 6791 (1993).

[33] B. Chen, Q.-M. Zhang, and J. Bernholc, Phys. Rev. B **49**, 2985 (1994).

[34] P. Bogusławski, E. L. Briggs, and J. Bernholc, Appl. Phys. Lett. **69**, 233 (1996).

[35] J.A. Van Vechten, J.D. Zook, R.D. Hornig, and B. Goldenberg, Jpn. J. Appl. Phys. **31**, 3662 (1992).

[36] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68**, 1829 (1996).

[37] C. H. Park and D. J. Chadi, Phys. Rev. B **55**, 12995 (1997).

[38] C. H. Park nad D. J. Chadi, Appl. Phys. Lett. **66**, 3167 (1995).

94

Table I. Effects of atomic relaxations around neutral substitutional impurities: Δb is the change of the bond length, E_{rel} is the relaxation energy, and ϵ_{imp} is the position of the impurity level, with e.m. denoting effective-mass states.

	Δb (%)	E_{rel} (eV)	ϵ_{imp} (eV)
	GaN		
C_{Ga}	−18.1	1.65	e.m.
Si_{Ga}	−5.6	0.65	e.m.
Ge_{Ga}	−1.4	0.25	e.m.
C_N	−2.0	0.1	$E_v+0.2$
Si_N	13.6	3.9	$E_v+1.2$
Ge_N	13.5	4.1	$E_v+1.35$
	AlN		
C_{Al}		unstable	
Si_{Al}	−3.0	0.3	e.m.
Ge_{Al}	17.2	0.9	$E_c-1.0$
C_N	2.0	0.9	$E_v+0.5$
Si_N	17.5	6.9	$E_v+2.0$
Ge_N	17.5	7.4	$E_v+2.1$

Table II. Energies in eV of substitutional (denoted by subst.) impurities on the cation site, and of DX configurations relative to those of the ground state (GS); and distances between the impurity and the host N^* atom. Unstable configurations are denoted by unst. See text for notation.

dopant	subst.	b_1(Å)	DX_1	b_1(Å)	DX_2	b_2(Å)
			GaN			
C(+)	0.05	1.62	0.65	3.05	GS	2.85
C(0)	0.05	1.63	0.60	3.25	GS	2.96
C(−)	0.1	1.62	0.60	3.31	GS	3.15
Si(0)	GS	1.85	unst.		unst.	
Si(−)	GS	1.85	unst.		unst.	
Ge(0)	GS	1.93	unst.		unst.	
Ge(−)	GS	1.93	0.60	2.61	unst.	
			AlN			
C(+)	GS	1.62	0.15	2.91	0.30	2.89
C(0)	unst.		0.05	2.97	GS	3.05
C(−)	unst.		0.7	2.97	GS	3.38
Si(+)	GS	1.85				
Si(0)	GS	1.87	0.3	2.82	unst.	
Si(−)	1.4	1.88	GS	2.98	0.95	2.93
Ge(+)	GS	2.07	unst.		unst.	
Ge(0)	GS	2.26	unst.		unst.	
Ge(−)	unst.		GS	2.90	unst.	

95

Table III. Calculated formation energies (in eV) of neutral substitutional impurities at cation-rich and N-rich conditions.

dopant	cation-rich	N-rich
GaN		
C_{Ga}	5.7	4.0
C_N	1.1	2.8
Si_{Ga}	0.9	1.4
Si_N	3.0	6.9
Ge_{Ga}	2.3	2.2
Ge_N	3.1	6.4
AlN		
C_{Al}	7.7	3.4
C_N	0.5	3.8
Si_{Al}	3.5	2.3
Si_N	2.6	10.0
Ge_{Al}	5.1	2.4
Ge_N	2.8	8.7

Table IV. Binding energies E_{pair} and short-range interaction energies E_{sr} for nearest-neighbor impurity pairs.

	E_{pair} (eV)	E_{sr} (eV)
GaN		
C-C	-1.1	-0.2
Si-Si	-0.8	-0.15
Ge-Ge	-1.1	-0.4
AlN		
C-C	-1.4	-0.5
Si-Si	-1.3	-0.7
Ge-Ge	-1.7	-1.0

Figure captions

Fig. 1. Atomic configurations for the non-relaxed substitutional impurity, and for DX states of C and Si. Light gray spheres – Ga, small black – N, and medium gray – impurity atoms, respectively.

Fig. 2. Schematic configuration diagram of an impurity, comparing energies of substitutional and DX-like states: (a) the DX state is unstable; in (b) and (c) both substitutional and DX states are local minima separated by an energy barrier; and in (d) the substitutional configuration unstable.

Fig. 3. Energies of DX_1 states of (a) Si, and (b) Ge impurities in negative charge states relative to the substitutional configurations. Circles, diamonds, triangles, and squares denote the variants with 3, 2, 1, and no Ga neighbors of N^*.

97

ideal

GaN:C(DX2)

b1

b2

AlN:Si(DX1)

GaN:C(DX1)
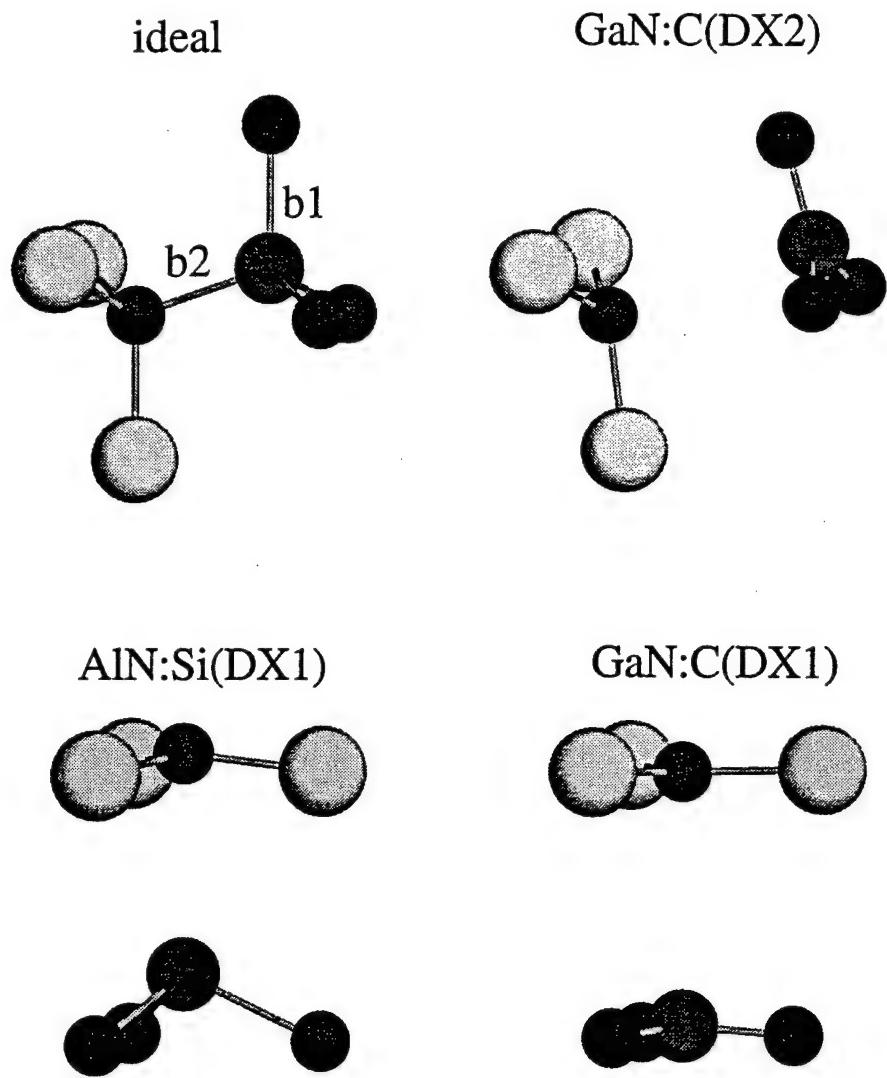
Figure 1

Boguslawski et al
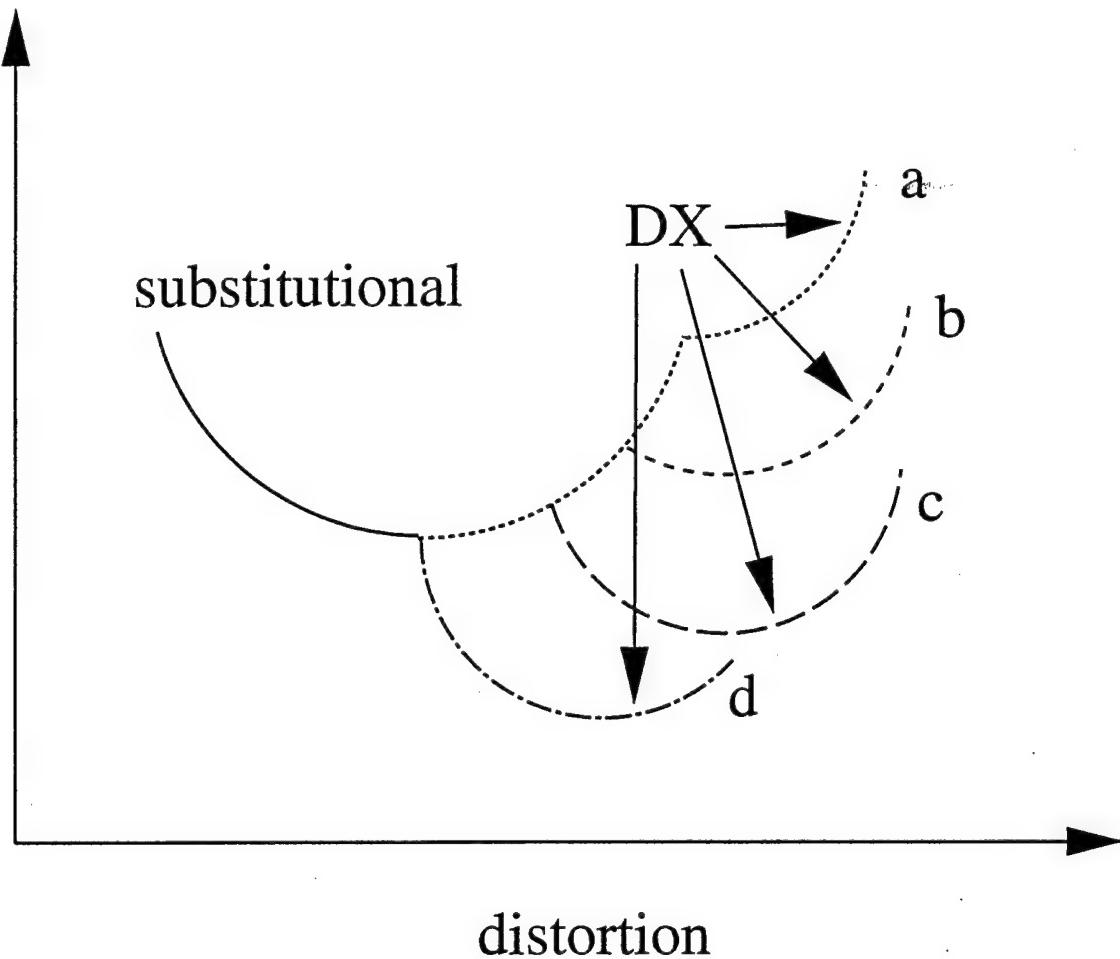
Figure 2

Boguslawski et al

(a) AlGaN : Si
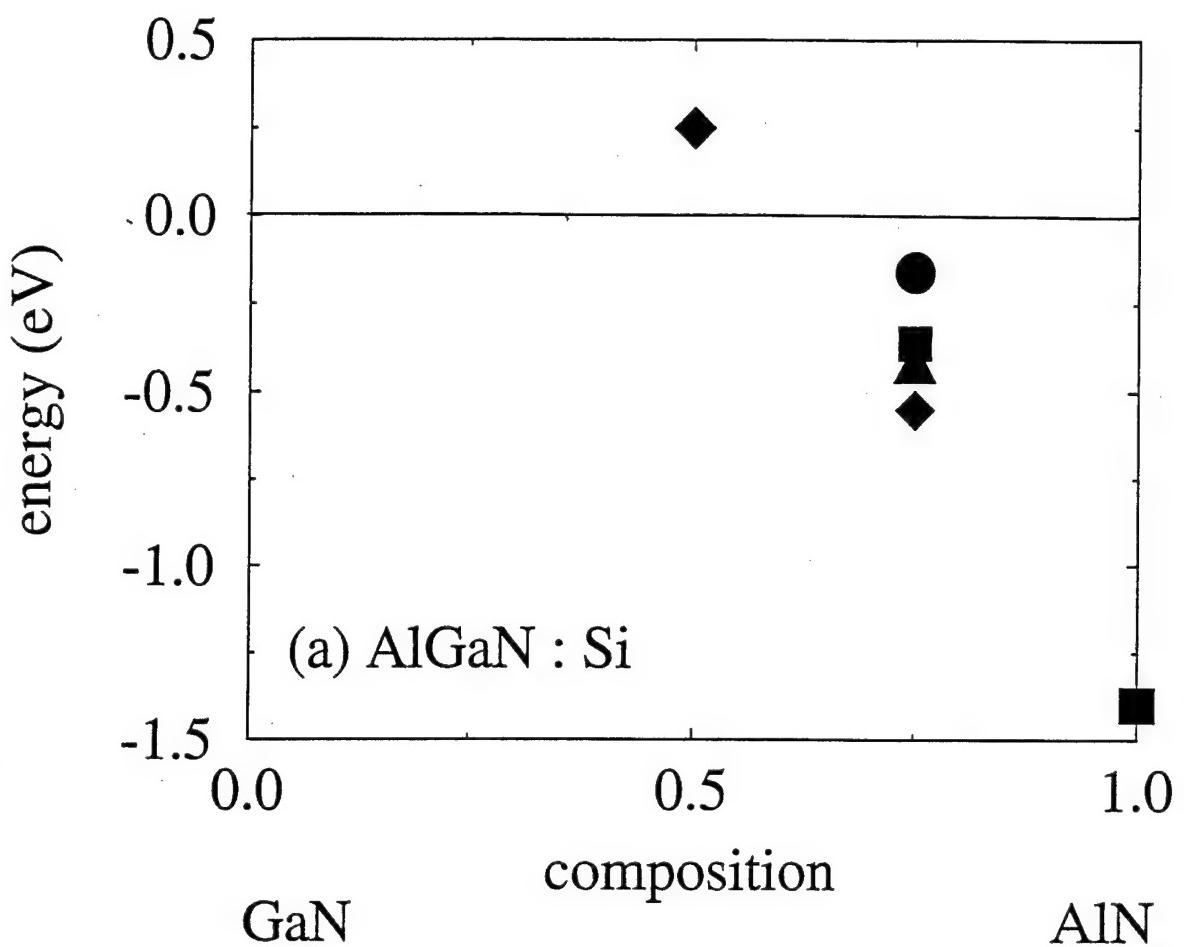
Figure 3a

Boguslawski et al

(b) AlGaN : Ge
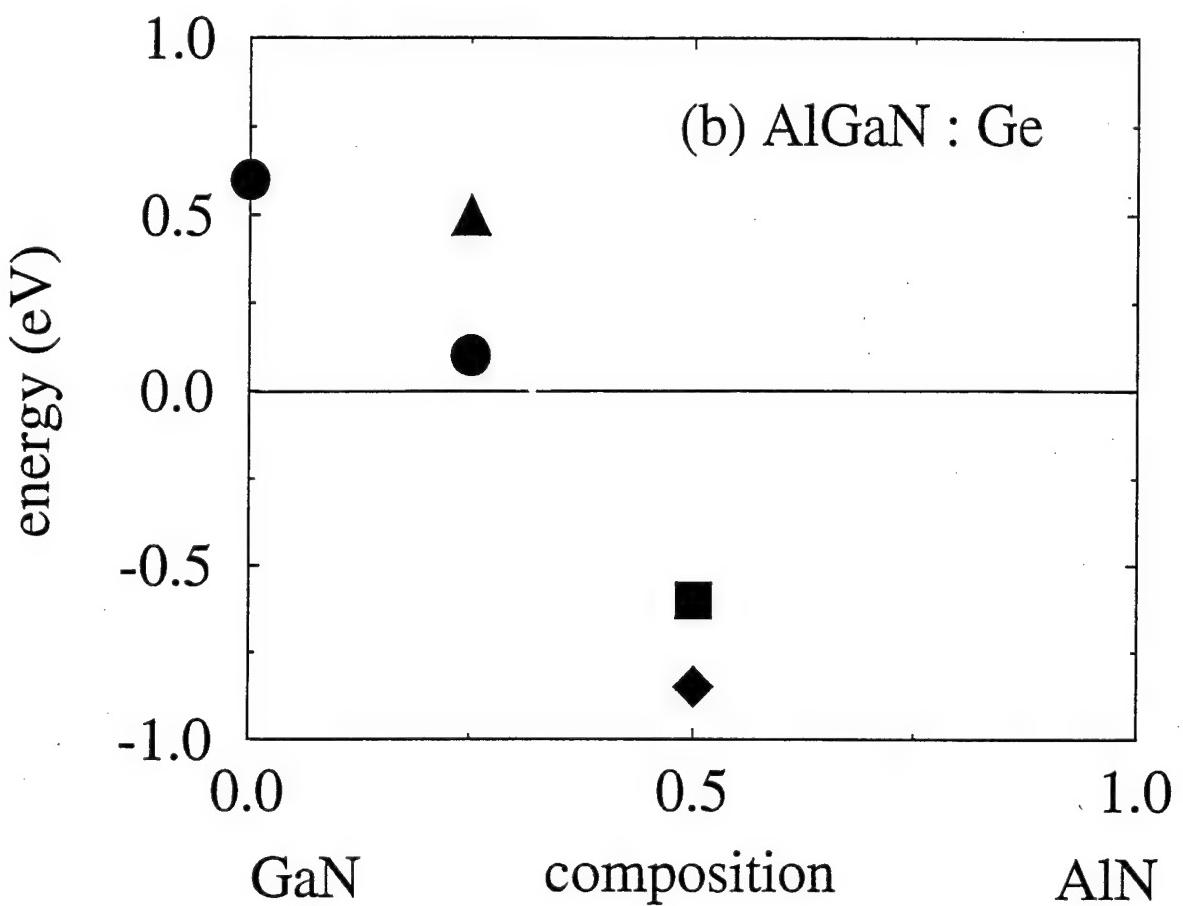
Figure 3b

Boguslawski et al

VIII. Diluent Gas Effects on Properties of AlN and GaN Thin Films Grown by Metalorganic Vapor Phase Epitaxy on α(6H)-SiC Substrates

Andrew Hanser, Colin Wolden[†], William Perry, Tsvetanka Zheleva, Eric Carlson, Philip Hartlieb, Robert F. Davis

Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC 27695
[†]Department of Chemical Engineering, Colorado School of Mines, Golden, CO, 80401

Abstract

Thin films of AlN and GaN were deposited on α(6H)-SiC(0001) wafers using metalorganic vapor phase epitaxy (MOVPE) and H_2 and N_2 diluents. A computational fluid dynamic model of the deposition process was used to analyze the film growth conditions for both diluents. Low temperature (12 K) photoluminescence of the GaN films grown in N_2 had peak intensities and full widths at half maximum of ~7 meV which were equal to or better than those films grown in H_2. Cross-sectional and plan-view transmission electron microscopy of GaN films grown in both diluents showed similar microstructures with a typical dislocation density of $10^9/cm^2$. Hall measurements of n-type (Si doped) GaN grown in N_2 revealed Hall mobilities equivalent to those films grown in H_2. Acceptor-type behavior of Mg-doped GaN grown in N_2 was repeatably obtained without post-growth annealing. Secondary ion mass spectrometry revealed equivalent levels of H in Mg-doped GaN films grown in both diluents.

A. Introduction

Blue and green light emitting diodes based on III-Nitrides are commercially available,[1] and the fabrication and room temperature continuous wave operation of nitride-based short wavelength injection laser diodes with lifetimes over 10,000 hours have recently been reported.[2] Additionally, microelectronic devices for high-temperature, high-frequency, and high-power applications are being pursued.[3,4] Efforts are ongoing to further improve device performance and to better understand the factors which influence film properties.

In the following sections, the results of a computational fluid dynamics model in an inverted MOVPE reactor similar to that used in this research for the growth of III-Nitride materials are presented. The stagnation point flow reactor geometry approximated in our reactor has the potential to produce large area films that are uniform in both thickness and composition[5,6]. Under ideal conditions a flow field is obtained that results in uniform heat and mass transfer gradients across the deposition radius. However, the effects of buoyancy and finite geometries lead to the development of thermal recirculation flows that disrupt the ideal transport conditions. The model results include analysis of the predicted heat and mass transfer gradients and were applied to the deposition of AlN and GaN thin films on SiC substrates. The gas flow structure and the influence of the diluent gas on film properties were investigated. The diluent gas usually constitutes $\geq 50\%$ of the gas phase in the reactor. As such, it essentially helps determine the flow structure in the reactor and often affects the decomposition of the source species. The thermal conductivity and viscosity of the gas phase and the molecular diffusivity of species through this phase are also important in the MOVPE process. H_2 is the primary diluent used in the MOVPE deposition of films of AlN, GaN, and AlGaN alloys; however, it has been shown to have deleterious effects on the incorporation of In in indium-containing nitride films.[7] The effect of the diluent species, if any, on growth of AlN and GaN are not well known. The influence of H_2 and N_2 diluents on the optical, electrical, and microstructural properties of our AlN and GaN thin films is presented.

B. Experimental Procedures

Computer Modeling. A computational fluid dynamics model was developed using FIDAP, a commercial software package that employs the finite element method, to obtain a better understanding of the factors governing thin film growth in our MOVPE system The velocity and temperature fields were calculated by solving the following coupled conservation equations in cylindrical coordinates:

$$\nabla(\rho \upsilon) = 0 \qquad \text{Continuity Equation} \qquad (1)$$

$$\rho(\upsilon \bullet \nabla \upsilon) = -\nabla P + \mu \nabla^2 \upsilon + \rho g \qquad \text{Conservation of Momentum} \qquad (2)$$

$$\rho C_p (\upsilon \bullet \nabla T) = \nabla \bullet (k \nabla T) \qquad \text{Conservation of Energy} \qquad (3)$$

The simplified energy balance neglects contributions from Dufour effects, viscous dissipation, radiation, and heat generated by reactions. Previous work[6,8] has shown that it is safe to ignore these effects under low pressure MOVPE conditions. The calculations were performed for a binary mixture of ammonia and the diluent gas (H_2 or N_2). The temperature-dependence of the transport properties were abstracted from the Sandia database[9,10] and included in the simulations. The viscosity and thermal conductivity of the mixture were determined using the semiempirical formula developed by Wilke.[11] It is appropriate to neglect contributions of the metalorganic precursors as they constitute less than 0.01% of the mixture. Buoyancy effects were included in the momentum balance.

The reactor geometry shown in Fig. 1 was simulated in the computer model. The two viewports were excluded to make the system perfectly axisymmetric. The velocity boundary conditions included a zero velocity condition at the inlet and reactor walls. The velocity at the substrate and susceptor were fixed at the rotation speed of 30 RPM. The boundary conditions included neither a radial gradient at the centerline nor an axial gradient at the outlet.

A uniform axial velocity across the gas inlet was assumed. This velocity profile was achieved experimentally following the approach used to generate flat flames for combustion studies.[11] The reactants were first passed through a packed bed to ensure good mixing. A
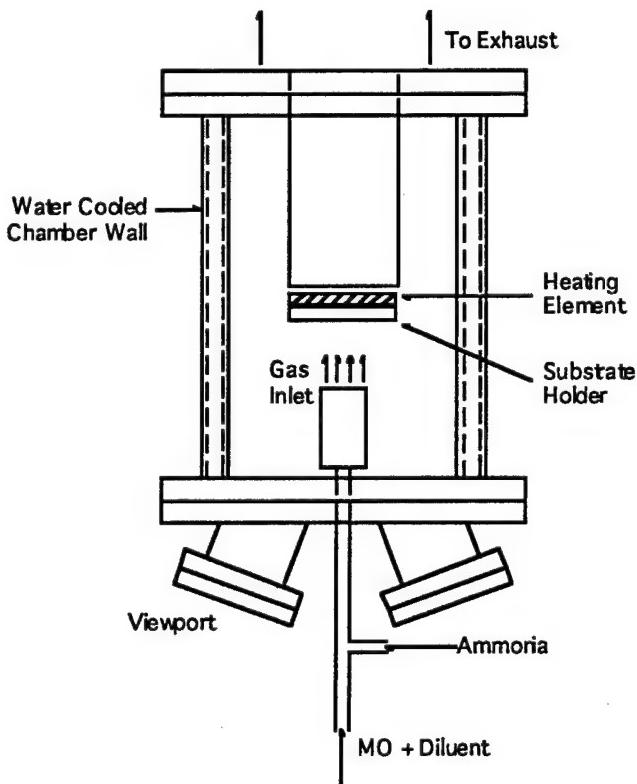
Figure 1.　　Schematic of the MOVPE deposition chamber

uniform velocity profile was subsequently created by flowing the reactants through a 1 inch long honeycomb structure created by a packed array of thin wall, 1 mm diameter silica tubes. The temperature boundary conditions included in the model were fixed temperatures at the inlet, reactor walls and susceptor and the absence of both a radial gradient at the centerline and an axial gradient at the outlet. A nonuniform mesh employing nine-node biquadratic elements was used to divide the reactor geometry into discrete elements. A typical mesh included ~2000 elements which produced ~20,000 equations and unknowns that were solved using a combination of successive substitution and Newton-Raphson techniques. All calculations were performed at the North Carolina Supercomputing Center.

Film Growth, Doping and Characterization. The AlN and GaN thin films were grown on on-axis 6H-SiC(0001) substrates in an inverted flow, vertical cold wall MOVPE reactor similar to the model described in the previous section and elsewhere.[12] The MOVPE system was operated under computer control using a LabVIEW-based program developed in-house. The substrate was heated to the deposition temperature under a diluent flow of either H_2 or N_2 at 2.2 standard liters per minute (slm). The deposition pressure was 45 Torr. The substrate was rotated continuously at approximately 30 RPM during deposition. Nitrogen was used as the carrier gas for the metalorganic precursors in all depositions. Trimethylaluminum (TMA) and triethylgallium (TEG) were the Group III sources, and ammonia (NH_3) was the nitrogen source. The ammonia was delivered to the deposition chamber through a line separate from the metalorganic precursors. Silane (SiH_4) diluted in H_2 was the source for the n-type Si dopant; bis-(cyclopentadienyl) magnesium (Cp_2Mg) was the source for the p-type Mg dopant.

A high temperature (1100 °C) AlN film approximately 1000Å thick was grown directly on each 6H-SiC(0001) substrate as the buffer layer. The GaN deposition conditions were optimized for each diluent. The flow rate of TEG in both diluent gases was 25.0 μmol/min and the flow rate of NH_3 was 1.6 slm. The SiH_4 was introduced into the reactor at flow rates between 0.15 and 3.75 nmol/min; the Cp_2Mg was introduced into the reactor at flow rates between 0.13 and 0.25 μmol/min. After terminating the GaN growth the substrate was cooled at a controlled rate to room temperature under the flowing diluent and NH_3.

The photoluminescence (PL) properties of the films were determined at 12 K using a 15 mW He-Cd laser (λ=325 nm) as the excitation source. Transmission electron microscopy (TEM) was performed at 200 keV using standard sample preparation and imaging techniques. Capacitance-voltage (CV) measurements were made using a mercury (Hg) probe and a computer controlled Hewlett-Packard 4284A LCR meter. Contacts for n-type Hall-effect measurements were made using annealed Ti/Au bilayers; contacts for p-type measurements were made using as-deposited and annealed Ni/Au bilayers. The contacts were deposited on the corners of 7mm × 7mm samples in the Van der Pauw geometry. The contacts were annealed in a RTA system under flowing N_2 at 600 °C for one minute. Hall effect measurements were

performed at room temperature. Secondary ion mass spectrometry (SIMS) was conducted using a Cameca IMS-6F system and GaN thin films implanted with H and Mg as standards.

C. Results and Discussion

Figures 2a and 2b show the calculated velocity contours and temperature contours obtained under the deposition conditions employed for our MOVPE reactor with the use of H_2 and N_2 diluents, respectively. A 60% diluent 40% NH_3 gas mixture was used in the modeling. Figure 2a shows the velocity contours for both diluents to be similar, with minimal recirculation cells forming below the substrate. Figure 2b demonstrates one of the differences between the two diluents. The higher thermal diffusivity of H_2 generates higher temperatures further away from the substrate, with a temperature boundary layer approximately 1.5 times as thick as with the N_2 diluent. Uniform temperature gradients are achieved across a majority of the substrate for both diluents. In the mass transport limited regime, the general conclusions can be extended from heat to mass transfer, resulting in predicted uniform transport of the growth species to the substrate across a large area of the substrate. This was experimentally verified using electron microscopy, where a film thickness uniformity variation of ~5% was observed over a 2.54 cm diameter.

The as-grown GaN films had highly reflective surfaces with no visible pits or cracks. A cross-sectional TEM micrograph of a 1 μm GaN film grown in H_2 is shown in Fig. 3. The
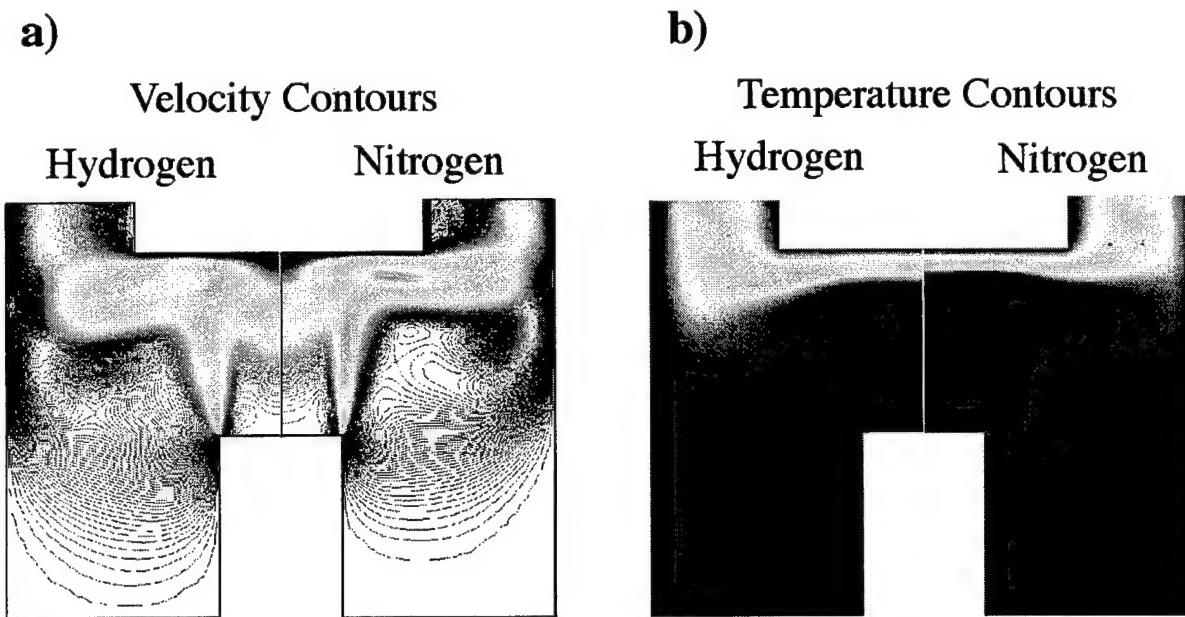
Figure 2.   a) velocity contours of the diluent/NH_3 fluid in the MOVPE reactor under deposition conditions. b) temperature contours of the diluent/NH_3 mixture under deposition conditions. Conditions for both figures were: $T_{substrate}$ = 1000°C, P_{system} = 45 Torr, T_{inlet} and T_{wall} = 25°C, V_{inlet} = 250 cm·sec^{-1}

GaN has a high concentration of defects near the AlN-GaN interface which decreases with distance from the interface. The microstructures of the films grown in both diluents were very similar.

Low temperature (12 K) PL measurements were made on GaN films grown in both diluents. Figure 4 shows the PL spectra for undoped GaN films grow in N_2 diluent. The spectra for films grown in N_2 were characterized by strong near band edge (NBE) emission with a FWHM of 6.90 meV and "yellow" emission levels that were three orders of magnitude lower than the NBE emission. The single feature of the NBE at 358.4 nm (3.460 eV) has been
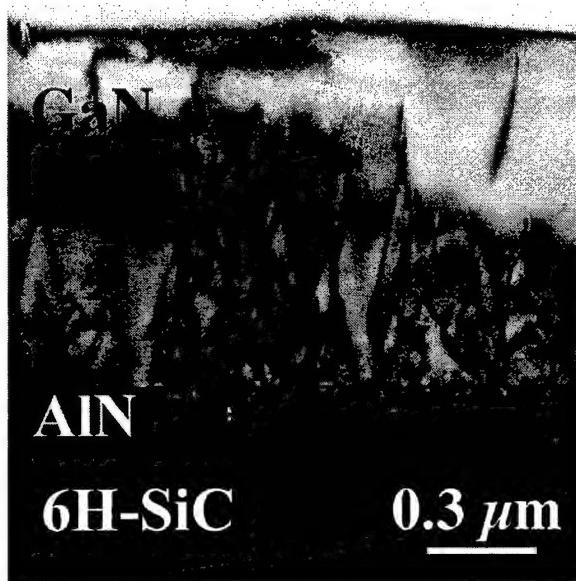
Figure 3.   Cross-sectional transmission electron micrograph of a GaN film grown on an AlN buffer layer. Plan view TEM measurements revealed a dislocation density of $\sim 10^9$ cm^{-2}.
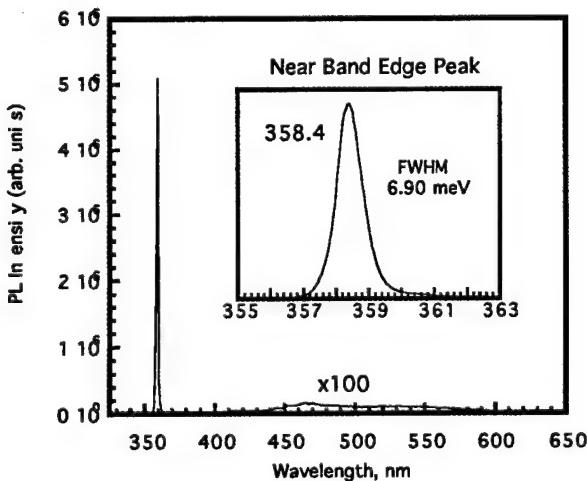
Figure 4.   12 K photoluminescence of GaN grown in a N_2 diluent.

107

attributed to an exciton bound to a neutral donor (I_2).[13] The PL spectra for H_2 diluent films were characterized by weaker NBE emission and stronger yellow emission levels, possibly due to non-optimized growth conditions.

Undoped GaN films grown in both H_2 and N_2 were too resistive for Hall measurements. Controlled n-type doping with Si was achieved from $\sim 5 \times 10^{16}$ to $\sim 5 \times 10^{18}$ cm^{-3}. The maximum room temperature Hall mobility for a 1 μm thick film was 275 cm^2/V·s at a carrier concentration of 1×10^{17} cm^{-3}. Magnesium doped GaN films deposited in a nitrogen diluent have repeatably shown p-type behavior as determined by CV measurements without post-growth annealing. All films grown in hydrogen required a post-growth thermal anneal to achieve p-type behavior. Hall measurements on as-grown Mg-doped GaN samples grown in N_2 revealed carrier concentrations as high as 2×10^{17} cm^{-3} and a Hall mobility of ~ 13 cm^2/V·s. The as-deposited resistivites of these films ranged between 5-6 Ω·cm.

The formation of Mg-H neutral complexes in GaN has been proposed as the passivation mechanism for Mg acceptors.[14,15] As such, the concentrations of H in undoped and Mg-doped GaN films grown in H_2 and N_2 diluents were investigated using SIMS. Figure 5 shows the H concentration as a function of dopant concentration in Mg-doped GaN films grown in N_2 and H_2 diluents. The H levels for Mg-doped films grown in both diluents were ~ 1.5 to 2 orders of magnitude higher than those in the undoped films. The H level of the film grown in the H_2
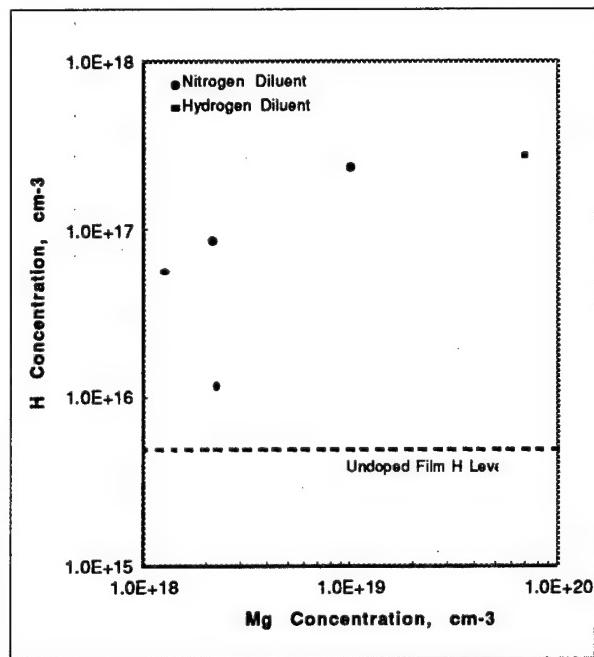
Figure 5.   Hydrogen concentration in Mg-doped GaN films vs. magnesium concentration for films grown in N_2 and H_2 diluents.

diluent was on the order of the H levels in the films grown in the N_2 diluent. Additionally, the H level for the film grown in H_2 was ~2 orders of magnitude lower than the Mg level in the film. This would suggest that there is not sufficient H to compensate all the Mg acceptors.

D. Conclusions

Metalorganic vapor phase epitaxy was used to deposit AlN and GaN thin films on on-axis α(6H)-SiC(0001) wafers using H_2 and N_2 diluents. A computational fluid dynamics computer model of the deposition process was employed to analyze the film growth conditions using both diluents. Low temperature (12 K) photoluminescence measurements of GaN films grown in N_2 had peak intensities and full widths at half maximum of ~7 meV which was equal to or better than those obtained for films grown in H_2. Transmission electron microscopy of films grown in both diluents showed similar microstructures. Room temperature Hall measurements of Si doped, n-type GaN grown in N_2 revealed Hall mobilities equivalent to those films grown in H_2, with a maximum value of 275 cm^2/V·s. Acceptor-type behavior of Mg-doped GaN grown in a N_2 diluent was repeatably obtained without post-growth annealing. Analysis via SIMS of Mg-doped films grown in H_2 and N_2 revealed H concentrations on the same order for both diluents, and ~2 orders of magnitude lower than the Mg doping level in the films.

E. Future Work

Future work include further analysis of the effect of the diluent on the as-grown acceptor-type behavior of Mg-doped GaN thin films. Further SIMS and PL analysis will be employed with rapid thermal annealing studies to obtain an explanation for the acceptor passivation mechanism. Growth of undoped and Si- and Mg-doped $Al_xGa_{1-x}N$ will be examined in both H_2 and N_2 diluents to compare doping ranges as a function of AlN content in the alloy. GaN and $Al_xGa_{1-x}N$ p-n junction devices, as well as $Al_xGa_{1-x}N$/GaN MESFET devices will be grown, fabricated and characterized.

F. Acknowledgments

G. References

1. S. Nakamura, M. Senoh, N. Iwasa, S. Nagahama, Jpn. J. Appl. Phys. **34**, L797 (1995).
2. S. Nakamura, 2nd Internat. Conf. on Nitride Semicond. Tokushima, Japan, Oct. 25-31, 1997.
3. Y.-F. Wu, S. Keller, P. Kozodoy, B. P. Keller, P. Parikh, D. Kapolnek, S. P. Denbaars, U. K. Mishra, IEEE Electron. Dev. Lett. **18**, 290 (1997).

4. O. Aktas, Z. F. Fan, A. Botchkarev, S. N. Mohammad, M. Roth, T. Jenkins, L. Kehias, H. Morkoç, IEEE Electron. Dev. Lett. **18**, 293 (1997).
5. C. Houtman, D. B. Graves, K. F. Jensen, J. Electrochem Soc. **133**, 961 (1986).
6. G. Evans, R. Grief, Trans. ASME **109**, 928, (1987).
7. E. L. Piner, M. K. Behbehani, N. A. El-Masry, F. G. McIntosh, J. C. Roberts. K. C. Boutros, S. M. Bedair, Appl. Phys. Lett. **70**, (4) 461.
8. J. P. Jenkinson, R. Pollard, J. Electrochem. Soc. **131**, 2911, (1984).
9. R. J. Kee, F. M. Rupley, J. A. Miller, Sandia National Laboratories Report SAND87-8215, (1994).
10. R. J. Kee, G. Dixon-Lewis, J. Warnatz, M. E. Coltrin, J. A. Miller, Sandia National Laboratories Report SAND86-8246, (1995).
11. C. R. Wilke, J. Chem. Phys. **18**, 517 (1950).
12. A. D. Hanser, C. A. Wolden, W. G. Perry, T. Zheleva, E. P. Carlson, A. D. Banks, R. J. Therrien, R. F. Davis, submitted to J. Electron. Mater.
13. R. Dingle, D. D. Sell, S. E. Stokowski, M. Ilegems, Phys. Rev. B **4**, 1211 (1971).
14. S. Nakamura, N. Iwasa, M. Senho, T. Mukai, Jpn. J. Appl. Phys. **31**, 1258 (1992).
15. W. Götz, N. M. Johnson, J. Walker, D. P. Bour, Appl. Phys. Lett. **67**, 2666 (1995).

IX. Optical Memory Effect in GaN Epitaxial Films

V. A. Joshkin, J. C. Roberts, F. G. McIntosh, and S. M. Bedair
Department of Electrical and Computer Engineering, North Carolina State University,
Raleigh, NC 27695-7911

E. L. Piner, M. K. Behbehani
Department of Materials Science and Engineering, North Carolina State University,
Raleigh, NC 27695-7907

Abstract

Memory effects are reported in the optical properties of GaN and AlN epitaxial films grown by atmospheric pressure metalorganic chemical vapor deposition. After exposing selected areas of particular samples with He-Cd laser light (3.8 eV), a persistent and marked decrease was observed in the near band edge photoluminescence (PL) intensity emitted from these areas. This effect has been observed in epitaxial films that typically have a pyramid-like hillock surface. This ability to modulate PL emission intensity at individual points in these materials can be exploited as a method for optical data storage. A means of erasing information stored using this effect has also been investigated using lower energy (~2 eV) photons.

A. Introduction

During the last 25 years, considerable effort has been devoted to the design of optical data storage devices [1-6]. Despite the commercial success of high-density, magneto-optical data storage systems, vigorous research activity continues regarding the development of non-magnetic, all-optical storage media. Most of these investigations are based on materials that trap electrons at deep defects. Read and write data transfer rates in electron trapping media should be fast because the process is photon-electronic rather than thermal in nature. One of the most promising technologies is based on electron trapping in alkaline earth crystals doped with rare earth elements [7-9]. A number of investigators have also studied the optical memory effect in AlN ceramics [10] and AlGaAs alloys [11]. In these technologies, information is written when photoionization of deep electron traps effectively sensitizes these materials by creating metastable states that modulate their electrical and/or optical properties. The written information can be retrieved from these sensitized crystals in various ways. For example, when the sensitized areas of the crystal are exposed to a "reading" laser beam, electrons can escape from the traps and produce photons with near band edge energy [7-10] or the beam can be diffracted by a locally varying refractive index due to variations in space charge [11]. Another effect has been observed related to electron trapping in epitaxial GaN and AlN that also has potential to be used for optical data storage systems.

The intensity of room temperature (RT) near band edge photoluminescence (PL) of GaN is significantly decreased at areas that have been exposed to a sufficient dose of ultraviolet (UV) radiation, and that this effect can be reversed i.e., "erased" with illumination of these areas with longer wavelength laser light.

The investigated films were grown by atmospheric pressure metalorganic chemical vapor deposition (MOCVD) at 950°C on (0001) sapphire substrates. Trimethylgallium (TMG, -10°C) and trimethylaluminum (TMA, +18°C) were used as the column III sources, ammonia (NH_3, 100%) was used as the column V source, and nitrogen was used as the carrier gas. Following the initial deposition of a ~100 A AlN buffer layer grown at 700°C by atomic layer epitaxy (ALE), a ~0.3 μm $Al_{0.1}Ga_{0.9}N$ lower cladding layer and the 1 μm GaN layer were grown at 950°C. The surfaces of these films have pyramid-like hillocks, that are similar to results of investigations of GaN grown by MOCVD in a nitrogen ambient reported by Sasaki *et al.* [12]. The resistivity of these films have been estimated to be ~1 Ω·cm and the films exhibit double crystal x-ray diffraction (DCXRD) rocking curves of the (0002) plane that have a full width at half maximum (FWHM) of ~100 arc sec. This optical memory effect in AlN films grown directly on sapphire by ALE at 700°C has been investigated.

The RT PL spectra of these films has been studied using a 10 mW He-Cd laser (325 nm), which was focused to a spot size ~100 μm in diameter. The He-Cd laser was used not only for PL excitation but also for longer UV exposures of the sample that induced the optical memory effect. A 4 mW He-Ne laser operating at 632.8 nm with a ~1 mm diameter spot size (unfocused) was used to reverse i.e., "erase" the optical memory effect that was induced by the He-Cd laser beam. The samples were mounted on a motorized linear translation stage. For studying the optical memory effect, PL measurements were typically performed by keeping the He-Cd laser beam focused at a fixed position and moving the sample through the beam at a controlled rate, thereby obtaining a spatial record of emitted PL at the near band edge transition energy (~3.4 eV for GaN). Conventional PL spectra i.e., intensity vs wavelength, can also be measured with this setup by keeping the sample stationary and scanning the monochromator over wavelength.

The RT PL spectra of the GaN films exhibited intense near band edge emission at 3.4 eV and very weak yellow emission at 2.3 eV. Prolonged UV excitation of some GaN samples resulted in a marked decrease of their PL emission intensity with time, and this effect persisted long after removal of the UV excitation. This memory effect was illustrated in curve a of Fig. 1, which showed the decrease of near band edge (~ 3.4 eV) PL emission intensity versus time of He-Cd laser excitation. Curve b of Fig. 1 illustrates the timing and duration of the sample excitation. The sample was initially illuminated for 5 minutes with the He-Cd laser, and the PL emission decreased appreciably, saturating at an intensity less than half of the initial measured intensity. The ~20 second delay of measuring the PL emission (curve a) from the

112

GaN sample is related to the stabilization time of the lock-in amplifier after the initial exposure of the sample to the He-Cd beam. After 5 minutes of irradiation, the He-Cd laser was switched off and the sample was allowed to "recover" at RT in ambient light. The He-Cd laser was switched on to make five more measurements over a three-day period to measure the PL intensity in the recovering film. The laser was turned on only long enough to obtain a stable reading of PL emission, and then immediately turned back off to minimize further degradation of PL emission intensity. After a three-day period the PL emission was still not fully recovered with respect to the initial measured intensity.

The long lifetime of the PL intensity recovery in these films constituted a memory effect that can be used to record information over the areal extent of the material. By exposing selected points to various doses of UV radiation, their subsequent emission to optical excitation can be modulated compared to unirradiated areas. Figure 2 illustrates a demonstration of this concept where 8 points, A - H, all known distances from one another and lying on the same line, have been irradiated by focused He-Cd laser light for durations of 10, 5, 10, 2, 2, 5, 1, and 1 min. respectively. A baseline measurement of PL emission (3.4 eV) was made along this line before irradiating these points, as shown in curve a, Fig. 2. It should be noted that points D and E are separated by ~50 μm, points G and H are separated by ~100 μm, while the other points have spatial separations of ~1 mm or more. Immediately following the UV exposure of these isolated points, the sample was scanned at a rate of ~1.5 mm/min from a location just before point A through point H and PL emission was collected at the near band edge energy of 3.4 eV. The results of this measurement are shown in curve b, Fig. 2 and indicate that the ~20 % decrease in PL emission intensity can be effectively "read" at each point of this scan. Subsequent PL scans made after a period of two and three days, as shown in Fig. 2 curves c and d respectively, demonstrated the persistence of this memory effect at RT. Since points G and H (~100 μm spacing) can be spatially resolved, while points D and E (~50 μm spacing) cannot, the spatial resolution was apparently limited by the He-Cd laser spot size (~100 μm). It should also be mentioned that when conventional PL spectra were measured at irradiated points A - H, no shift was observed in peak emission energy (3.4 eV) and no change in the FWHM of the PL spectra.

A preliminary investigation was made of whether or not longer wavelength radiation has an influence on this optical memory effect. First, a GaN sample (~1 μm thick) was written at two different points, separated by about 150 μm, by exposing the sample to He-Cd laser light. The PL intensity (measured at 3.4 eV) was measured along this linear region to confirm that the PL intensity was diminished at these two points, as shown in Fig. 3, curve a. Next, a 4 mW He-Ne laser (632.8 nm) with a 1 mm (defocused) spot size was used to simultaneously irradiate these two closely spaced points for 30 minutes. Within 3 minutes of the He-Ne
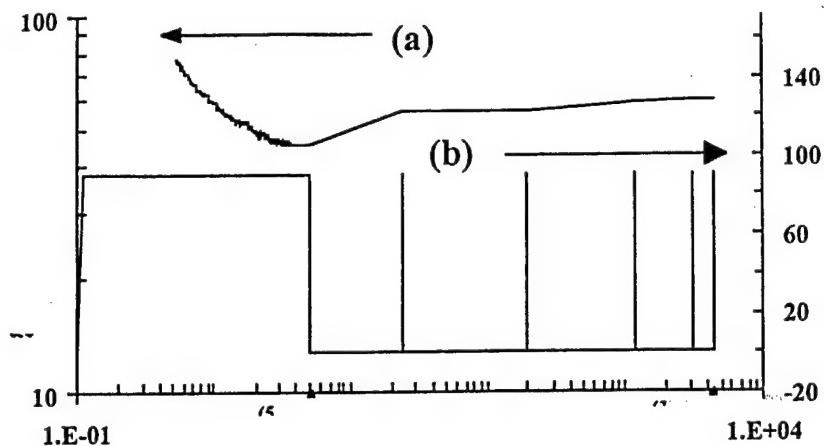
Time dependence of room
temperature PL intensity at

Figure 1. (a) Time dependence of PL emission intensity at 3.4 eV.
           (b) Timing and duration of He-Cd laser excitation for the PL measurement of
                (a).
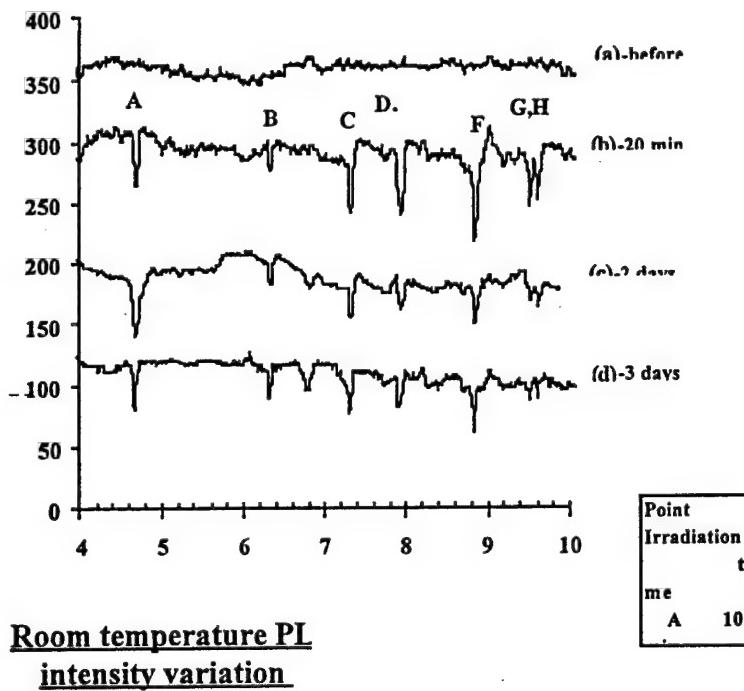
Room temperature PL
intensity variation

Figure 2. (a) PL emission intensity (@ 3.4 eV) vs. distance along the sample prior to
                UV "writing" exposure.
           (b) PL emission intensity (@ 3.4 eV) vs. distance along the sample
                immediately after "writing" at points A-H with He-Cd nm laser light for
                various durations.
           (c) PL scan similar to (b) made two days after writing at points A-H.
           (d) PL scan similar to (d) made three days after writing at points A-H.

114

Room temperature PL
intensity variation
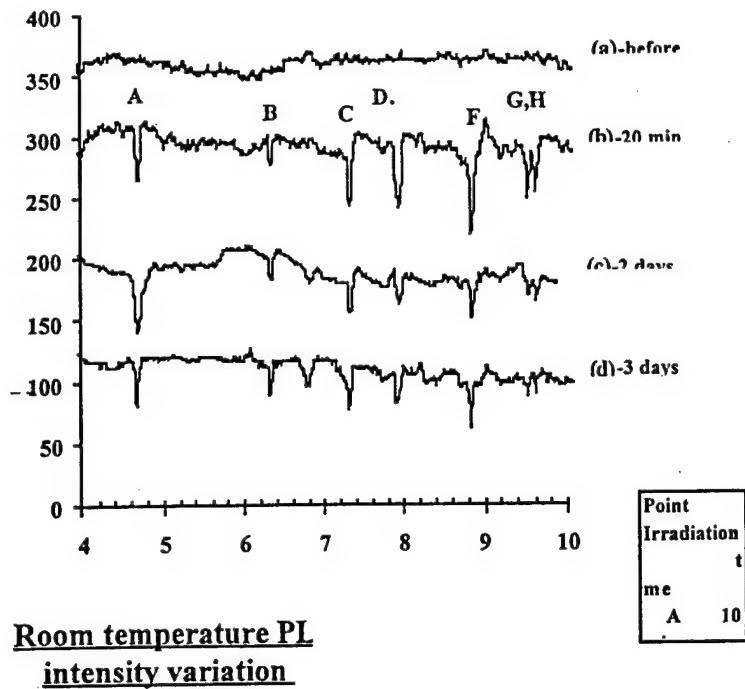
Figure 3.   PL emission intensity (@ 3.4 eV) vs. distance along sample:
(a) After "writing" at two points along the line with He-Cd laser light.
(b) Three minutes after irradiating the points with He-Ne laser light.
(c) Six hours after irradiating the points with He-Ne laser light.

irradiation, another linear scan of PL emission intensity was performed that exhibited little apparent change, as shown in Fig. 3, curve b. However, 6 hours after the He-Ne irradiation, a second linear scan of PL emission exhibited a nearly uniform intensity profile, shown in Fig. 3, curve c, indicating that the optical memory effect had been effectively erased.

This optical memory effect was observed in a thin AlN film grown by ALE at 700°C. Because the reduction in PL emission intensity with irradiation time was even more pronounced than in the aforementioned GaN films, a slight variation in the measurement method was used to measure the effect in the AlN film. To avoid overexposing the film and loose the PL signal completely, an initial scan was made by moving the AlN sample through the He-Cd beam at a rate of ~1.5 mm/min while acquiring the emission intensity vs wavelength data. This measurement approach limited the UV exposure of each point on the sample to approximately 9 seconds, assuming a laser spot size of 100 μm, and the resulting spectra is displayed in curve a of Fig. 4. Curve b of Fig. 4 shows the PL emission measured with the sample held at a fixed position during the 10 minute monochromator scan. While both spectra exhibit a deep level peak at ~3.1 eV, the second scan with the prolonged UV exposure is 30 times less intense than that of curve a. The yellow emission at 2.25 eV was not diminished by the prolonged UV exposure. It should be noted that similar deep level emission near 3.1 eV was been observed previously from AlN and was attributed to oxygen impurities [13].

115
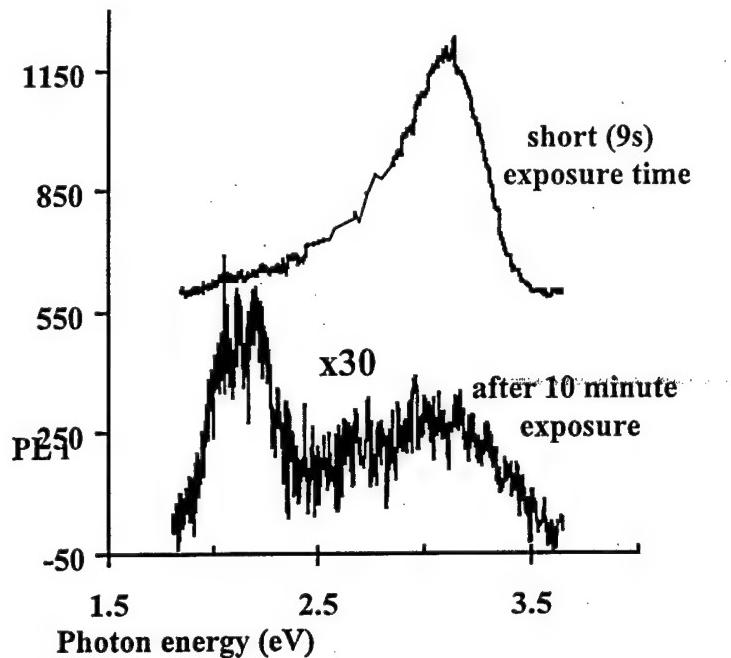
Figure 4.   PL intensity vs. wavelength of an ALE grown AlN film:
    (a) Data taken by translating sample at 1.5 mm/min. under the He-Cd laser
         beam (9 sec. of UV exposure per point).
    (b) "Standard" PL measurement of the same AlN film, where the total
         exposure time was 10 minutes during the scan at a single point on the film.

To explain the phenomena, it is suggested that the effect is related to donor-like states, probably due to oxygen. One of these states is the fundamental state near the conduction band of GaN. The other, charged metastable state (or states) with their accompanied large lattice distortion, can produce nonradiative recombination of excitons [14]. This metastable state can be created from the fundamental state by trapping of an electron during the He-Cd laser exposure. The "lifetime"of this metastable state appeared to be long, retaining its charge for days. The increasing of the density of metastable states during He-Cd laser exposure resulted in the reduction of near band edge emission as shown in Fig. 2. More work is needed to have a better understanding of not only the fundamental nature of these defect levels, but also the role that supposed oxygen impurities and strain have in their creation and their relaxation.

In conclusion, a long term optical memory effect was observed in GaN epitaxial films having structured, grainy surfaces. This effect was demonstrated by "writing" at spatially resolved points with UV laser energy and then "reading" these points by monitoring a decrease in near band edge PL emission compared to the unirradiated area between the points. Subsequent irradiation of written points with red light appeared to erase the optical memory effect after a sufficient delay. The potential applications of this phenomenon toward optical recording of information was highlighted.

References

1. R. A. Bartolini, H. A. Weakliem, and B. F. Williams, *Review and Analysis of Optical Recording Media*, Selected Papers on Optical Storage, ed. G. T. Sincerbox and J. M. Zavislan, SPIE Milestone Series **MS 42**, 488-497.
2. D. Smith, *Magnetic Films and Optics in Computer Memories*, Selected Papers on Optical Storage, ed. G. T. Sincerbox and J. M. Zavislan, SPIE Milestone Series **MS 42**, 528-547.
3. D. Psaltis and F. Mok, Scientific American, Nov. 1995, 70-76.
4. K. A. Rubin and M. Chen, Progress and Issues of Phase-change Erasable Optical Recording Media, Selected Papers on Optical Storage, ed. G. T. Sincerbox and J. M. Zavislan, SPIE Milestone Series **MS 42**, 621-629.
5. E. Udd, Optical Information Processing, ed. S. Jutamulia and J. Tsujiuchi, Part 2, Proc. IEEE **84**, 884 (1996).
6. S. Jutamulia, G. M. Storti, W. Seiderman and J. Lindmayer, Optical Data Storage Technologies, ed. S.-J. Chua and J. C. McCallum **1401**, 113 (1991).
7. J. Lindmayer, P. Goldsmith and K. Gross, Optical Data Storage Technologies, ed. S.-J. Chua and J. C. McCallum **1401**, 103 (1991).
8. D. T. Brower and R. E. Revay, Optical Data Storage, ed. D. B. Carlin and D. B. Kay **1663**, 867 (1992).
9. S. Chen, C. Qi, F. Dai and F. Gan, 3rd International Symposium on Optical Storage, ed. F. Gan, SPIE **2053** 114 (1993).
10. J. H. Harris and R. A. Youngman, Wide Band Gap Semicondctors, ed. T. D. Moustakas, J. I. Pankove and Y. Hamakawa, Mat. Res. Soc. Symp. Proc. **242**, 451 (1992).
11. P. Mortensen, Laser Focus World, p. 38, Aug. 1995.
12. T. Sasaki and T. Matsuoka, J. Appl. Phys. **77**, 192 (1995).
13. J. H. Harris and R. A. Youngman, Photoluminescence and Cathodoluminescence of AlN, Properties of Group III Nitrides, ed. J. H. Edgar, EMIS Datareviews Series No. 11, P. 203 (1994).
14. K. J. Malloy and K. Khachaturyan, *Imperfections in III/V Materials* (ed. E. R. Weber), Semiconductors and Semimetals **38**, 235 (1993).

X. Growth and Photoluminescence Properties of $In_xGa_{1-x}N$ Films Grown on SiC Substrates by Low Pressure Metalorganic Vapor Phase Epitaxy

A. D. Hanser, W. G. Perry, R. F. Davis
Department of Materials Science and Engineering, North Carolina State University
Raleigh, NC 27613-7907

Abstract

The microstructures and photoluminescence (PL) spectra have been determined for $In_xGa_{1-x}N$ films (x ≤ ~0.23) grown on substrates of α(6H)-SiC(0001) wafer/AlN buffer layer/GaN heterostructures by low pressure metalorganic vapor phase epitaxy at 780°C using nitrogen as the diluent and carrier gas and V/III ratios as low as 2,420. Indium droplets were not observed. The InN content in these films was limited by the deposition pressure in the system. The maximum InN content achievable at 45 Torr was ~13%. Increasing the deposition pressure to 90 Torr increased the maximum InN content to ~23%. Room temperature and 12 K PL spectra of the films revealed single-feature, near band edge (NBE) emission with increasing full width at half maximum (FWHM) values with increasing In fraction. The PL NBE FWHM for an $In_{0.23}Ga_{0.77}N$ film at 12 K was 103 meV.

A. Introduction

Gallium nitride (GaN) and the related alloys $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ are the current materials of choice for the fabrication of light emitting diodes in the blue and green regions of the visible spectrum [1] and laser diodes in the blue and violet regions [2]. In these devices, $In_xGa_{1-x}N$ is the active layer and the wavelength of the emitted light is controlled by the InN content in the film. The weak In-N bond at the surface and the resultant high equilibrium vapor pressure of nitrogen [3,4] as well as the tendency to phase separate [5] are the main difficulties in growing In-containing materials. Films of $In_xGa_{1-x}N$ with relatively high InN content (> 10%) and without In droplets have recently been obtained by using high In source flow rates, high V/III ratios and lower growth temperatures [6-8]. The use of N_2 as the carrier and diluent gas instead of H_2 has also been shown to increase the InN incorporation in the films [9,10]. Metalorganic vapor phase epitaxy (MOVPE) at one atmosphere of pressure has been the primary method of growth for the $In_xGa_{1-x}N$ materials. Relatively little research using this technique for growth of these materials at reduced pressure has been conducted [11,12]. In this letter, the growth and photoluminescence (PL) properties of $In_xGa_{1-x}N$ films grown by low pressure MOVPE are reported.

B. Experimental Procedures

All films were grown on on-axis α(6H)-SiC(0001) substrates using a MOVPE reactor described elsewhere [13]. A high temperature (1100°C) AlN film approximately 1000Å thick was grown directly on the substrate as a buffer layer. A 1 μm thick GaN film was then deposited as a template for the $In_xGa_{1-x}N$ films. The latter films were grown for 30 minutes at 780°C using trimethylindium (TMI) and triethylgallium (TEG) as the In and Ga sources, respectively. The TEG flow rate was 3.8 μmol/min and the TMI flow rate was varied between 5 and 30 μmol/min. N_2 was used as the carrier and diluent gas. The NH_3 flow rate was 0.082 mol/min (2 standard liters per minute). The growth experiments were conducted at 45 and 90 Torr. The films were characterized via scanning electron microscopy (SEM) using a JEOL 6400 FE microscope operating at 5 keV. Room temperature and low temperature (12 K) photoluminescence (PL) measurements were made using a 15 mW He-Cd laser (λ=325 nm) as the excitation source. All film compositions were determined from PL peak positions and compared to published compositional values.

C. Results and Discussion

The growth rate of the $In_xGa_{1-x}N$ films was approximately 2000Å/hr as measured by cross-sectional SEM. The surfaces of films with an InN content < ~10% were smooth with only a few pits for film thicknesses up to 1000Å, the largest thickness observed in this research. As the InN content increased over ~10% and the thicknesses of these films increased to over several hundred angstroms, a pitted and grooved morphology was observed, as shown

119

in Fig. 1. This surface morphology could be attributed to either strain in the film or a change in the growth mode for higher InN content films. The lattice parameter difference between GaN and InN (Δ/a_0) is 11.2% [14]. Assuming no phase separation, as the InN content of the $In_xGa_{1-x}N$ film increases the strain in the film increases due to increasing lattice mismatch between it and the underlying GaN layer. This increase in strain can lead to a change in the growth mode from two-dimensional (Frank-van der Merve) to three-dimensional growth (Stranski-Krastanov). With this latter growth mode, the coalescence of islands leads to layer formation and can result in the observed pitted and grooved surface morphology.

Under no growth conditions were In droplets observed. Previous research [8,15] has indicated the necessity of using V/III ratios in the range of 5,200–10,000 to eliminate the presence of In droplets; however, $In_xGa_{1-x}N$ films with approximately 23% InN content were grown in this research with a V/III ratio of approximately 2,420.

The InN content of the films was a function of reactor pressure during growth. Figure 2 shows the near band edge (NBE) emission peak positions at 12 K for $In_xGa_{1-x}N$ films grown under varying TMI flow rates as a function of reactor pressure. At 45 Torr the In incorporation in the film increased linearly with increasing TMI flow rate from 5 to 15 μmol/min. However, from 20 to 30 μmol/min the InN content remained constant at approximately 13%. At a reactor pressure of 90 Torr, the $In_xGa_{1-x}N$ NBE emission increases linearly with increasing TMI flow rate up to 30 μmol/min. This difference in In incorporation is due to the higher overpressure of gas at the growth surface at 90 Torr which reduces the ability of the In to desorb from
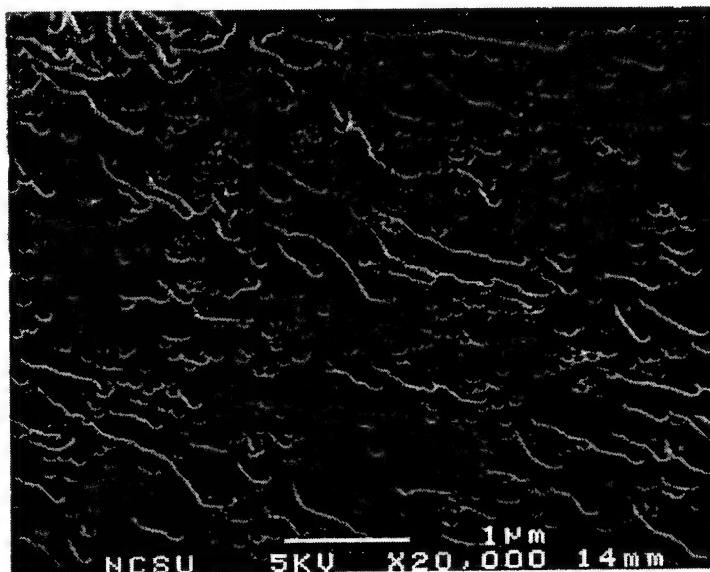
Figure 1.   SEM micrograph of the surface of an 1000Å $In_{0.23}Ga_{0.77}N$ film.

120

the surface. The rate of desorption of the source species from the growth surface during MOVPE growth is a kinetic process which can be represented by Eq. 1:

$$r_d = k_d \times n_d \tag{1}$$

where r_d is the rate of desorption, k_d is the rate constant, and n_d is the number of atoms of evaporating species per unit area. The rate constant k_d is the fraction of these atoms evaporating per unit time and is a function of the temperature and pressure of the system. As the pressure is increased at a constant temperature, k_d will decrease and the number of atoms desorbing from the surface will be reduced. The larger number of gas phase collisions at the $In_xGa_{1-x}N$ growth surface due to the increased system pressure at 90 Torr decreases the number of In atoms desorbing from the surface, thus allowing more of this species to be incorporated during growth.

Figure 3 shows the 12 K PL of GaN and $In_xGa_{1-x}N$ films as a function of x from approximately 0.05 to 0.23. The peak intensities for all the films were on the same order of intensity. The compositions of the $In_xGa_{1-x}N$ films were determined using a least squares linear fit to data points from twelve published papers quoting $In_xGa_{1-x}N$ compositional values. The FWHM of the band edge related peaks increased with increasing In fraction, possibly due to an increase in the number of free carriers [12] or an increase in local variations of the film composition. The FWHM of the NBE peak for the PL spectra corresponding to an $In_{0.05}Ga_{0.95}N$ film centered at 372.4 nm (3.330 eV) was 52 meV; the analogous value for an $In_{0.23}Ga_{0.77}N$ film centered at 428.9 nm (2.891 eV) was 103 meV. The FWHM of the reference GaN NBE peak was 6.90 meV. No donor-to-acceptor pair (DAP) recombination was observed at 12 K. The NBE peaks showed some asymmetry toward the lower energy side, possibly due to compositional fluctuations in the films.
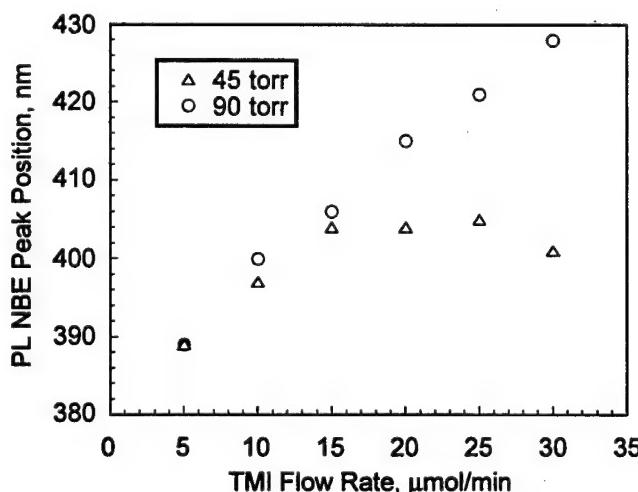
Figure 2.   The dependence of the photoluminescence peak position at 12 K of $In_xGa_{1-x}N$ films on MOVPE growth pressure and TMI flow rate.

The room temperature (RT) and 12 K PL spectra for an $In_{0.23}Ga_{0.77}N$ film are shown in Figs. 4 a) and b), respectively. At RT the peak position is at 433.9 nm (2.858 eV) and the FWHM is 174 meV. The deep level "yellow" emission in the film is weak relative to the NBE emission and no DAP recombinations are observed. At 12 K, the peak position shifts to 428.9 nm (2.891 eV) and the FWHM decreases to 103 meV. The broader peak at RT is most likely due to a higher concentration of free carriers. The 33 meV NBE peak shift for the $In_xGa_{1-x}N$ film is less than that of GaN (65.8 meV) over the same temperature range [16].

D. Conclusions

The growth of $In_xGa_{1-x}N$ films with $x \leq {\sim}0.23$ was achieved on substrates of 6H-SiC(0001)/AlN buffer layer/GaN heterostructures by low pressure MOVPE with V/III ratios as low as 2,420. Indium droplets were not observed. The InN content in these films was a function of the deposition pressure in the system. Room temperature and 12 K PL spectra showed single-feature NBE emission without DAP recombinations. The FWHM of these spectra increased with increasing indium fraction. The FWHM for an $In_{0.23}Ga_{0.77}N$ film at 12 K was 103 meV.

E. Future Work

Future work includes deposition of $In_xGa_{1-x}N$ alloys with higher InN content through an investigation of additional growth parameters, including growth temperature and TEG flow rate. A TEM study of the defect structure in $In_xGa_{1-x}N$ bulk films is currently underway. Single and multiple $In_xGa_{1-x}N$ quantum well structures, as well as $In_xGa_{1-x}N$/GaN superlattices will be fabricated to investigate their photoluminescence and structural properties. Blue LED structures with $In_xGa_{1-x}N$ active regions will be fabricated, and the electroluminescence properties of the devices will be determined.
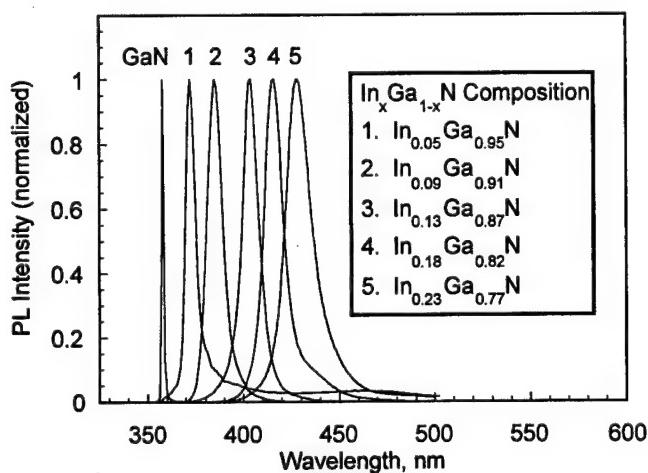
Figure 3.   Photoluminescence at 12 K of GaN and $In_xGa_{1-x}N$ films.
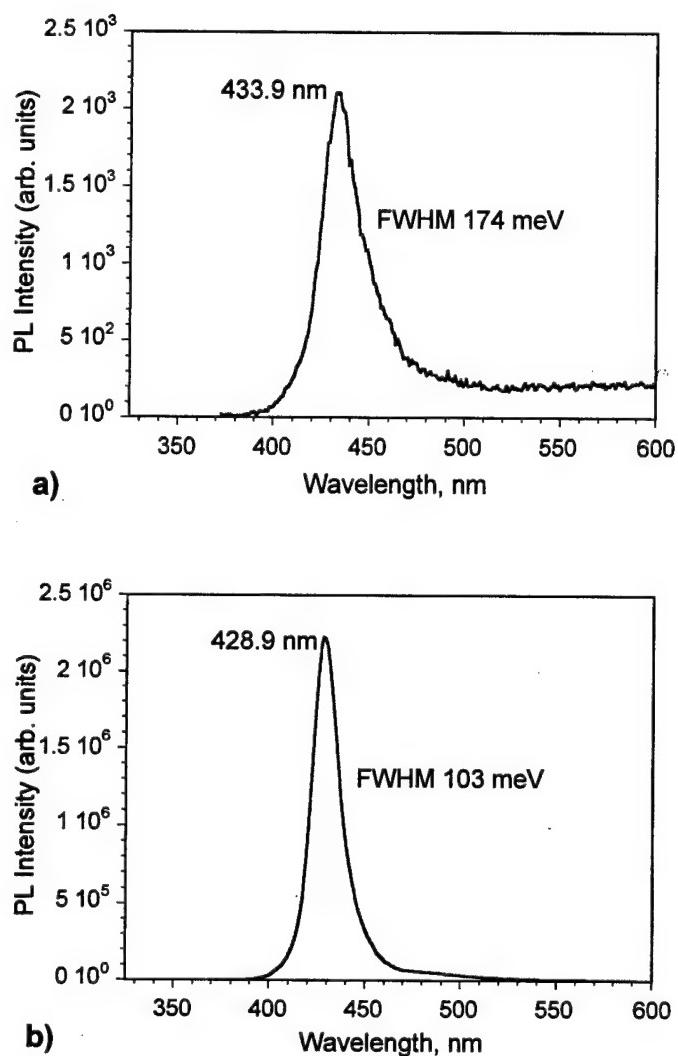
Figure 4. a) Room temperature and b) 12 K photoluminescence spectra of an In$_{0.23}$Ga$_{0.77}$N film. Note the different intensity scales for the two spectra.

F. Acknowledgments

G. References

1. S. Nakamura, M. Senoh, N. Iwasa, S. Nagahama, Jpn. J. Appl. Phys. **34**, L797 (1995).
2. S. Nakamura, J. Cryst. Growth **170**, 11 (1997).
3. R. D. Jones, K. Rose, J. Phys. Chem. Solids **48**, 587 (1987).
4. Q Guo, O. Kato, J. Appl. Phys., **73**, 7969 (1993).
5. I. Ho, G. B. Stringfellow, Appl. Phys. Lett. **69**, 2701 (1996).
6. S. Nakamura, T. Mukai, Jpn. J. Appl. Phys. **31**, L1457 (1992).

7. M. Shimizu, K. Hiramatsu, N. Sawaki, J. Cryst. Growth **145**, 209 (1994).
8. S. Keller, B. P. Keller, D. Kapolnek, U. K. Mishra, S. P. DenBaars, I. K. Shmagin, R. M. Kolbas, S. Krishnakutty, J. Cryst. Growth **170**, 349 (1997).
9. M. V. Weckwerth, K. P Killeen, R. Biefeld, T. Drummond, M. H. Crawford, J. C. Zolper, Electronic Materials Conference, Santa Barbara, June 1996.
10. E. L. Piner, M. K. Behbehani, N. A. El-Masry, F. G. McIntosh, J. C. Roberts. K. C. Boutros, S. M. Bedair, Appl. Phys. Lett. **70**, 461 (1997).
11. N. Yoshimoto, T. Matsuoka, T. Sasaki, A. Katsui, Appl. Phys. Lett. **59**, 2251 (1991).
12. T. Yuzhen, Z. Guoyi, J. Zhijian, D. Xiaozhong, W. Shumin, Solid State Comm. **102**, 405 (1997).
13. A. D. Hanser, C. A. Wolden, W. G. Perry, T. Zheleva, E. P. Carlson, A. D. Banks, R. J. Therrien, R. F. Davis, submitted to J. Electron. Mater.
14. *Properties of Group III Nitrides*, J. H. Edgar, ed., INSPEC, London, (1994).
15. W. Van der Stricht, I. Moerman, P. Demeester, L. Considine, E. J. Thrush, J. A. Crawley, Internet J. Nitride Semicond. Res. **2**, 16 (1997).
16. W. Shan, T. J. Schmidt, X. H. Yang, S. J. Hwang, J. J. Song, B. Goldberg, Appl. Phys. Lett. **66**, 985 (1995).

124

XI. Detection and Analysis of Phase Separation in Metalorganic Chemical Vapor Deposition InGaN

E.L. PINER*, N.A. EL-MASRY*, S.X. LIU*, and S.M. BEDAIR**
*Materials Science and Engineering Department, North Carolina State University, Raleigh, NC 27695
**Electrical and Computer Engineering Department, North Carolina State University, Raleigh, NC 27695

Abstract

InGaN films in the 0-50% InN composition range have been analyzed for the occurrence of phase separation. The 0.3 to 0.5 μm thick InGaN films were grown by metalorganic chemical vapor deposition (MOCVD) in the 690 to 780°C temperature range and were analyzed by θ-2θ x-ray diffraction (XRD), transmission electron microscopy (TEM), and selected area diffraction (SAD). As-grown films with up to 21% InN were single phase. However, for films with 28% InN and higher, the samples showed a spinodally decomposed microstructure as confirmed by TEM and extra spots in SAD patterns that corresponded to multiphase InGaN. An explanation of the data based on the GaN-InN pseudo-binary phase diagram is discussed.

A. Introduction

The GaN material system has been the focus of much research and development over the past several years, particularly since the realization of III-Nitride based optoelectronic devices emitting in the blue to violet spectrum. A primary achievement in the race to commercialize nitride semiconductors has been the growth of high quality InGaN used as the active layers in these high performance devices. However, in this race for efficient light emitters, some questions concerning the fundamental aspects of this material system have not been answered. The MOCVD growth of InGaN is quite complicated, especially for higher concentrations of In in the ternary film (>20%), leading to the unfortunate situation that many of its growth mechanisms are not understood. Obviously, a better understanding of these growth mechanisms will enhance the crystal quality as well as the reproducibility during production. In this article, the issue of phase separation in the InGaN ternary system is addressed.

Phase separation was first observed in InGaN by K. Osamura et al.[1] in 1975 while examining the x-ray diffraction patterns of annealed polycrystalline samples in the 600 to 700°C temperature range. More recently, R. Singh et al. reported experimental evidence of phase separation in molecular beam epitaxy grown InGaN using x-ray diffraction and optical absorption measurements.[2] Phase separation or quantum dots were also observed in $In_{0.20}Ga_{0.80}N$ (3 nm)/ $In_{0.05}Ga_{0.95}N$ (6 nm) multiple quantum well structures[3] using TEM and energy dispersive spectrometry (EDS). It was found that within the well region of these strained films the In composition varied in the range of 8-20%. On the theoretical side, the

large difference in interatomic spacing between GaN and InN (~11%) can give rise to a solid phase miscibility gap. Ho and Stringfellow[4] studied the temperature dependence of the binodal and spinodal lines in the InGaN system using a modified valence-force-field model to predict that at a growth temperature of 800°C, the equilibrium solubility of In in GaN would be less than 6%. In this study, bulk InGaN films were grown by MOCVD in the 0-50% InN alloy composition range and analyzed using θ-2θ XRD, TEM and SAD.

B. Experimental Procedure

InGaN films were grown on (0001) sapphire substrates at atmospheric pressure using a specially designed reactor that has been described elsewhere.[5,6] The pre-growth treatment includes annealing under nitrogen and passivating under ammonia at 1050°C for 15 and 1 minute, respectively. An AlN buffer layer grown by atomic layer epitaxy (ALE), deposited at 700°C and 100 Torr, was followed by an MOCVD grown AlGaN graded to GaN prelayer at 950°C and 750 Torr for 15 and 10 minutes, respectively, with a 2.5 minute grading period in between. The details of the growth conditions for the buffer layer are described elsewhere.[7,8] Trimethylgallium, trimethylaluminum, and ethyldimethylindium were used as the precursors with nitrogen as the carrier gas and hydrogen injected to control In incorporation. The InGaN layer was grown for 1 hour by MOCVD in the 690 to 780°C temperature range resulting in a film thickness of 0.3 to 0.5 μm. The concentration of In in the films was controlled by varying the growth temperature and the hydrogen flow rate.[7]

C. Results and Discussion

Figure 1 shows a series of θ-2θ XRD patterns with InN% increasing from (a) to (d). The In concentrations were determined by calculating the splitting between the (0002) InGaN peak position and the (0002) GaN peak position and applying Vegard's law. The samples were grown using identical growth conditions except the growth temperature of the InGaN epilayer was steadily decreased from (a) to (d) resulting in a steady increase in the InN% as indicated in the figure.[9] Diffraction patterns (a), (b), and (c), corresponding to 21, 28 and 40% InN in InGaN, respectively, are typical diffraction profiles for all samples with 40% or less

In content and indicate single crystal InGaN without In-metal on the surface. However, when the In content increases beyond 40%, an additional peak appears which is indicated as the "extra" peak in Figure 1(d) along with a significant increase of the count rate above background in the region between the (0002) InGaN and (0002) GaN peaks. Gaussian distribution functions were used to deconvolve the data in Figure 1(d) with the result that the extra peaks corresponded to 95, 36, and 14% InGaN as previously reported.[10] The 95% InN peak is represented by the "extra" peak while the 36 and 14% InN peaks are related to the high count region in Figure 1(d). The origin of these additional peaks will be discussed later.
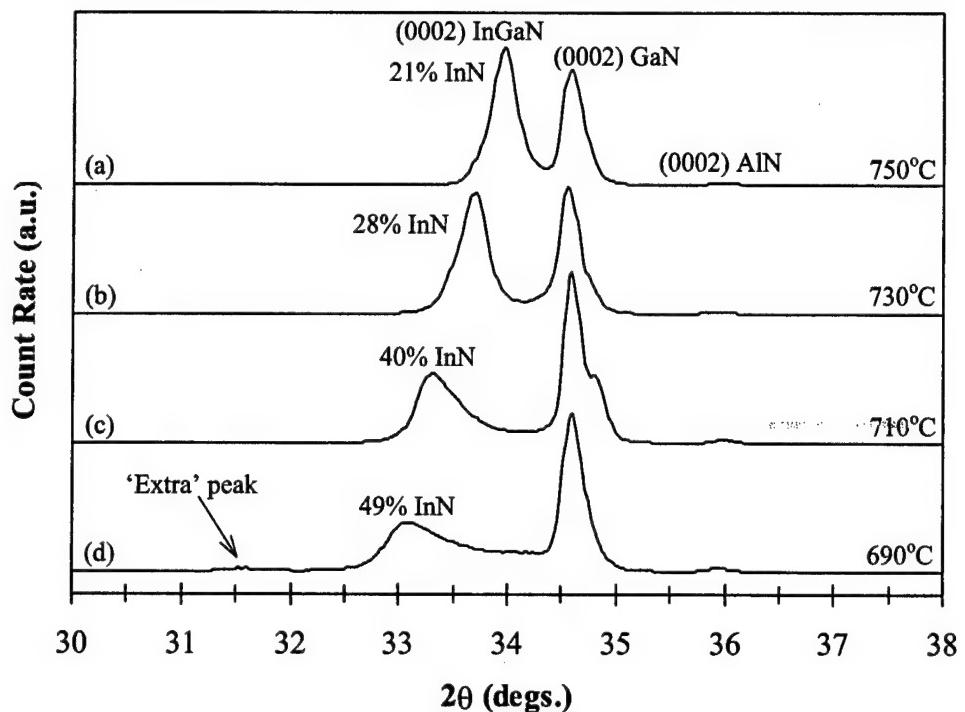
Figure 1.   θ-2θ XRD patterns of bulk InGaN epilayers grown at the growth temperatures and resulting in the In-content indicated in the figure. Note the appearance of the 'extra' peak at 31.44° in the film grown at 690°C.
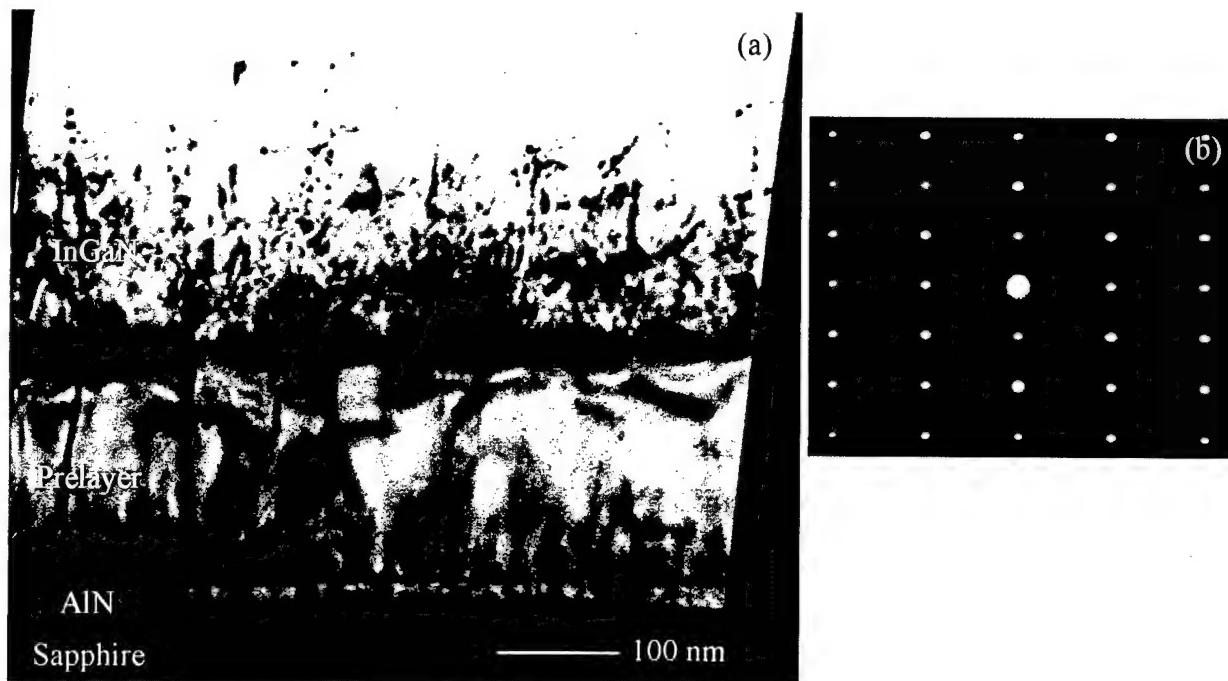

Figure 2.   (a) Dark field (g=c-axis) cross-sectional TEM image of InGaN with 21% InN. (b) [2-1-10] SAD pattern of the InGaN layer in (a). Selected area aperture revealing only the InGaN layer was used to obtain the diffraction pattern.

Fig. 2(a) shows a typical cross-sectional micrograph of a single phase InGaN film grown at 750°C with 21% InN. The micrograph was taken under two beam conditions with c-axis as the operating g-vector. Fig. 3(a) shows a microstructure of a multiphase InGaN film. Figs. 2(a) and 3(a) are micrographs of the films whose θ–2θ XRD patterns are represented in Figs. 1(a) and (b), respectively. Tilting the samples under the electron beam changes the dark contrast from region to region indicating the presence of strain in the films. The SAD patterns of the InGaN films are single crystalline as indicated by the absence of rings as shown in Figs. 2(b), and 3(b). The SAD pattern of the 21% InN film shown in Fig. 2(b) shows only a single set of diffraction spots indicating the single phase nature of this film. Fig. 3(b) shows a SAD pattern of a film with multiphase microstructure in which the diffraction spots exhibit a split which is pronounced in higher angle spots. The spots seem to split into multiple satellite spots, one which is weaker and is indicated by the arrow in Fig. 3(b). The diffraction pattern is indicating the single crystalline quality of the multiphase regions of the InGaN bulk film. The splitting of the diffraction spots confirms the presence of at least two phases in the microstructure. The θ-2θ XRD pattern of the multiphase InGaN film shown in Fig. 3(b) did not show additional peaks indicative of phase separation similar to Fig. 1(d). This discrepancy is due to the sensitivity of the SAD and XRD characterization techniques. For samples with less than or equal to 40% InN,
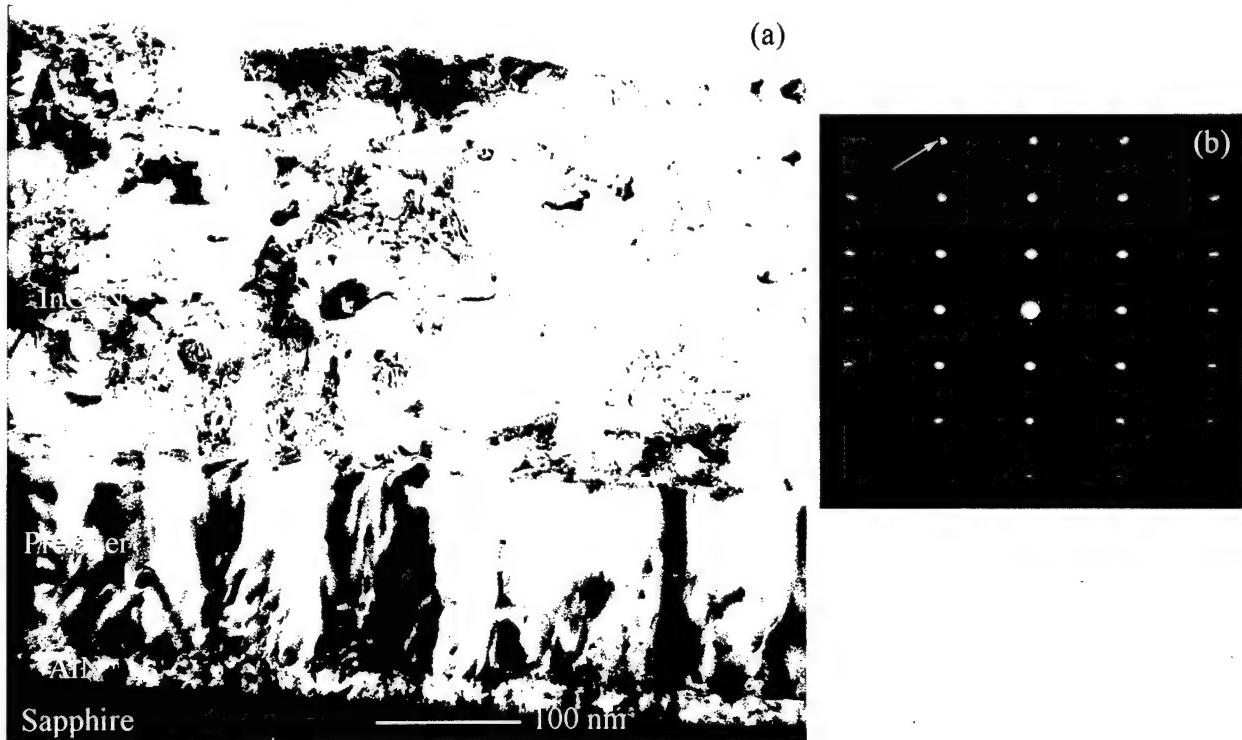
Figure 3. (a) Dark field (g=c-axis) cross-sectional TEM image of InGaN with 28% InN. (b) [01-10] SAD pattern of the InGaN layer in (a). Selected area aperture revealing only the InGaN layer was used to obtain the diffraction pattern.

the volume of the phase separated material is too small to be detected by XRD while SAD is sensitive enough to detect this small volume. However, for samples with greater than 40% InN, the volume of phase separated material has become sufficiently large to be detectable by XRD as shown in Fig. 1(d).

Cross-sectional TEM microstructural analysis of the 49% InN sample reveals a "tweed" appearance[11] at higher magnification in the areas with no strain contrast. This tweed structure is identified in Fig. 4 as the regions in which there appears alternating light and dark contrast fluctuations due to low and high In-content, respectively. This microstructural feature is characteristic of spinodal decomposition, thus, indicating its presence in the GaN-InN pseudo-binary system. The black regions in the micrograph are highly strained regions. A particularly well defined spinodally decomposed region in the micrograph is indicated by the arrow in the lower left corner. EDS in the STEM mode of the microscope was also used to characterize these films. However, the 100 Å diameter probe spot size made it impossible to distinguish between the separate regions of modulated composition due to their small size.

Although the theoretical equilibrium solubility of In in GaN has been calculated to be 6% at 800°C it should also be noted that, at this growth temperature, the spinodal composition is 22% In in GaN.[4] Figure 5 shows the GaN - InN pseudo-binary phase diagram determined by Ho and Stringfellow[4] with our experimental data superimposed. In the figure, the triangles and squares represent samples which are single phase and multiphase, respectively, as determine by SAD. Within the metastable limits of the spinode, represented by the dashed
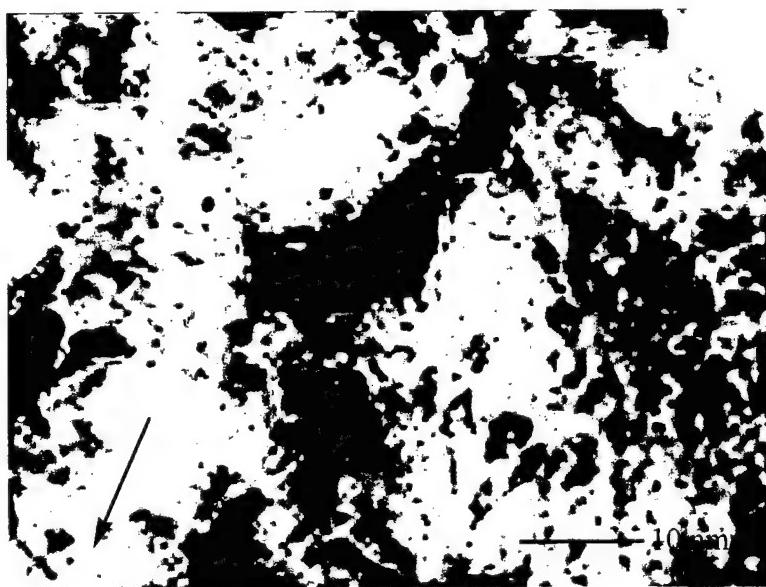
Figure 4.   Dark field (g=c-axis) cross-sectional TEM image of InGaN with 49% InN. Note the 'tweed' appearance indicative of spinodal decomposition in the lighter regions. The dark regions are due to high strain.

129

curve in Fig. 5, there is no barrier to nucleation and, therefore, the system is unstable because small fluctuations in composition that produce Ga-rich and In-rich regions will cause the total free energy to decrease. However, if the alloy composition lies between the spinodal limit and the miscibility gap (dashed and solid curves, respectively) small variations in composition lead to an increase in free energy and the system is no longer unstable but rather metastable and requires a nucleation and growth type transformation. Figure 5 indicates that the bulk grown samples with less than 21% InN does not exhibit phase separation and are, therefore, single phase. This value is within the range of being between the spinodal curve and miscibility gap, taking into account experimental scatter[12], and thus represent films which are metastable. We observe phase separation in samples with compositions of 28% InN and greater by SAD as indicated by the squares in figure 5. In this region of the GaN - InN pseudo-binary phase diagram the alloy is unstable and 'up-hill' diffusion can occur until the equilibrium compositions are reached which were calculated to be $In_{0.06}Ga_{0.94}N$ and $In_{0.92}Ga_{0.08}N$ at 800°C. These calculated equilibrium compositions are very close to the $In_{0.14}Ga_{0.86}N$ and $In_{0.95}Ga_{0.05}N$
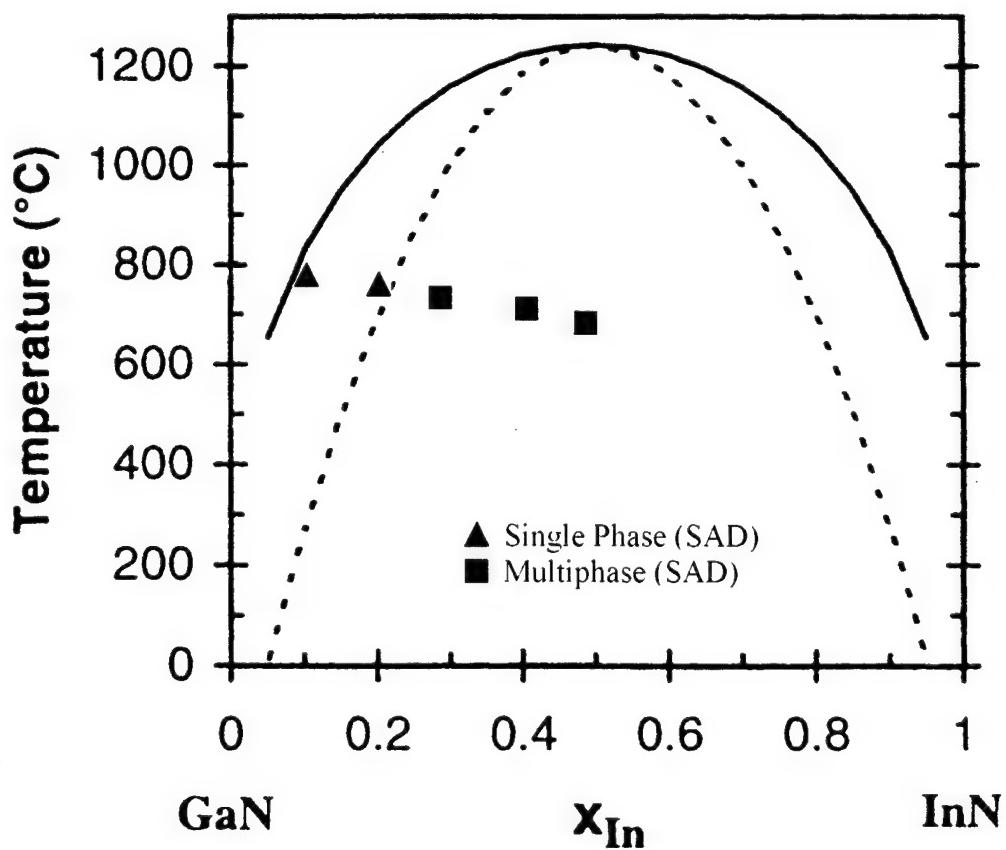
Figure 5.   Binodal (solid) and spinodal (dashed) curves for the InGaN system (Ref. 4).
             Experimental data is represented by triangles (single phase by SAD) and
             squares (multiphase by SAD) in the figure.

130

compositions determined from figure 1(d) discussed earlier thus offering validity to the thermodynamic calculations. The $In_{0.95}Ga_{0.05}N$ and $In_{0.14}Ga_{0.86}N$ peaks indicate the possibility of phase separation in the film while the $In_{0.49}Ga_{0.51}N$ and $In_{0.36}Ga_{0.64}N$ peaks represent non-equilibrium phases which can be a result of the MOCVD growth process. It is well known that vapor phase growth techniques are non-equilibrium and can allow the growth of alloys with compositions exceeding the solubility limits.[13]

D. Conclusions

We have demonstrated the presence of phase separation using both macroscopic (XRD) and microscopic (TEM and SAD) characterization techniques in MOCVD grown bulk InGaN. A transition from single phase to multiphase InGaN occurs for as-grown samples between 21 and 28% InN in InGaN indicated by additional spots in SAD. A tweed appearance is observed due to spinodal decomposition in the GaN - InN pseudo-binary system. The sensitivity of XRD is much less than that of SAD, which was able to show multiple spots for samples with 28% InN in InGaN whereas the observation of additional peaks in XRD did not occur until grater than 40% InN in InGaN. These results agree with the calculated spinodal decomposition range by Ho and Stringfellow[4] and can be useful in determining the limits of single phase growth by MOCVD.

E. Acknowledgments

F. References

1. K. Osamura, S. Naka, and Y. Murakami, J. Appl. Phys. **46**, 3432 (1975).
2. R. Singh, D. Doppalapudi, T. D. Moustakas, and L. T. Romano, Appl. Phys. Lett. **70**, 1089 (1997).
3. Y. Narukawa, Y. Kawakami, M. Funato, S. Fujita, S. Fujita, and S. Nakamura, Appl. Phys. Lett. **70**, 981 (1997).
4. I. Ho and G. B. Stringfellow, Appl. Phys. Lett. **69**, 2701 (1996).
5. K. S. Boutros, F. G. McIntosh, J. C. Roberts, S. M. Bedair, E. L. Piner, and N. A. El-Masry, Appl. Phys. Lett. **67**, 1797 (1995).
6. N. Karam, T. Parados, W. Rowland, J. Schetzina, N. El-Masry, and S. M. Bedair, Appl. Phys. Lett. **67**, 94 (1995).
7. E. L. Piner, M. K. Behbehani, N. A. El-Masry, F. G. McIntosh, J. C. Roberts, and S. M. Bedair, Appl. Phys. Lett. **70**, 461 (1997).
8. E. L. Piner, Y. W. He, K. S. Boutros, F. G. McIntosh, J. C. Roberts, S. M. Bedair, and N. A. El-Masry, Mater. Res. Soc. Symp. Proc. **395**, 307 (1996).

9. For an explanation of this phenomena consult the following reference: S. M. Bedair, F. G. McIntosh, J. C. Roberts, E. L. Piner, K. S. Boutros, and N. A. El-Masry, J. Cryst. Growth **178**, 32 (1997).
10. N. A. El-Masry, E. L. Piner, S. X. Liu, and S. M. Bedair, Appl. Phys. Lett. (Accepted for publication, ~ January 5, 1998).
11. D. A. Porter and K. E. Easterling, *Phase Transformations in Metals and Alloys* (Alden Press, Oxford, 1981), pp. 308-314.
12. G. B. Stringfellow, J. Cryst. Growth **27**, 21 (1974).
13. See, for example, G. B. Stringfellw, *Organometallic Vapor-phase Epitaxy: Theory and Practice* (Academic Press, San Diego, 1989).

XII. Selection, Growth, and Characterization of Gate Insulators on MOCVD Gallium Nitride for Use in High Power Field Effect Devices

R.J. Therrien*[†], O.H. Nam*, M.D. Bremser*, K. Linthicum*, H. Nimii*, E.P. Carlson*, G. Lucovsky**, R.F. Davis*

*Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC
27695-7907
**Department of Physics, North Carolina State University, Raleigh, NC 27695-8202

Abstract

Metal-insulator-semiconductor (MIS) capacitors have been fabricated on n-type GaN (0001) films using thermally grown Ga_2O_3, remote plasma enhanced chemical vapor deposited (RPECVD) SiO_2, and molecular beam epitaxy (MBE) AlN as the gate insulator and Al as the gate electrode. Each GaN epitaxial layer was grown by organometallic chemical vapor deposition (OMCVD) on a 6H-SiC(0001) substrate on which was previously deposited a 1000Å buffer layer of AlN. Nitrogen-free polycrystalline films of Ga_2O_3 were grown on the GaN. Capacitance-voltage measurements of capacitors fabricated from this oxide showed distinct depletion and accumulation regions with significant leakage. The AlN and SiO_2 capacitors demonstrated better electrical characteristics than the Ga_2O_3 because of lower leakage. The RPECVD SiO_2/GaN heterostructures, in particular, showed good agreement with the curves calculated for an ideal oxide and a small amount of hysteresis.

A. Introduction

Gallium nitride has physical and electronic characteristics which make it suitable for high temperature and high power devices.[1] Metal-insulator-semiconductor-field-effect-transistors (MISFETs) are of particular interest for these applications. As such, a high quality gate insulator is required. Metal-oxide-semiconductor (MOS) capacitors, fabricated using remote plasma deposited SiO_2 on n-type GaN, reportedly have low concentrations of both interface traps and fixed charge at the interface.[2] A MISFET device has also been fabricated using Si_3N_4, deposited via plasma enhanced chemical vapor deposition, as the gate insulator. This device exhibited a transconductance of 16 mS/mm at 30°C and 11 mS/mm at 200°C.[3]

Native oxides, grown by consuming the substrate in an oxidizing atmosphere, are the best insulators in many field effect devices. A native oxide, identified by x-ray diffraction as monoclinic β-Ga_2O_3, has been grown on GaN at temperatures above 750°C.[4]

Investigations regarding the efficacy of three insulating materials, namely, Ga_2O_3, AlN, and SiO_2, as gate insulators in a GaN-based field effect transistor are detailed in the following sections. Table I shows the physical and electrical properties of these insulators. Capacitance-voltage (C-V) measurements of the MIS capacitors fabricated on the GaN were used to determine their viability in FETs.

B. Experimental Procedure

Silicon-doped, n-type ($n \approx 1-2 \times 10^{17}$ cm^{-3}), 1–1.5 μm thick GaN epitaxial layers were grown via OMCVD at 1050°C and 40 Torr on a 1000Å AlN buffer layers previously grown at 1100°C on on-axis 6H-SiC (0001) substrates. Mercury probe C-V measurements verified the doping level. Details regarding the growth and doping procedures have been previously reported.[5]

The gallium oxide films were grown on the GaN films in a closed tube furnace. Prior to growth each GaN film/substrate assembly was cleaned *ex situ* sequentially in trichloroethlyene, acetone, and methanol and submersed in a 1:1 HCl:DI water dip. Residual carbon, chlorine, and oxygen remained on the surface, as shown by the Auger electron spectroscopy (AES) results in Fig. 1. The GaN was then placed on a quartz boat, inserted into a quartz tube furnace which was evacuated and backfilled with oxygen flowing through the system at 2500 sccm. The oxidation of the GaN surface was investigated for times ranging from three–six hours at 700°C, 800°C, 825°C, 900°C, 1000°C. The films were analyzed using electron dispersion spectroscopy (EDS) for composition and scanning electron microscopy (SEM) to observe the morphology.

The AlN films were deposited on the GaN layers by molecular beam epitaxy. The surface of each GaN sample was cleaned prior to deposition *ex situ* in the same manner as prior to the oxide growth, and *in situ* at 800°C in a 5 sccm ammonia flux for 30 minutes. All traces of oxide were removed using the latter procedure, as determined by x-ray photoelectron

Table I. Properties of the Selected Insulators

	Ga_2O_3	AlN	SiO_2
Bandgap (eV @ 300 K)	?	6.3	9
Dielectric Strength (10^7 V/cm)	?	0.6 - 1.5	1
Dielectric Constant	3-6	8.5	3.9
Thermal Conductivity W/cm·K	?	2.0	.014
Density (g/cm^3)	6.44	3.2	2.2

spectroscopy (XPS). Details for cleaning and surface preparation have been reported in Ref. 6. The AlN was then deposited at a temperature of 800°C for 20 minutes with an ammonia flux of 10 sccm.

The SiO_2 deposition was performed in a RPECVD system. Due to the inability to reach high temperatures with ammonia in this system, cleaning steps involving a nitrogen plasma were implemented. The GaN samples were cleaned with the aforementioned solvents and the HCl:DI water dip as before, placed upon a Si wafer in the RPECVD system, heated between 250° and 300°C and immersed in a nitrogen plasma. Varying plasma powers and system pressures were used in an attempt to remove all of the surface contaminates. A plasma power of 30W at a pressure of 0.1 Torr for 5 minutes removed all of the surface contaminates except oxygen, as shown in Fig. 2. The oxide was deposited on the GaN, following a 15 second exposure to an oxygen plasma using silane and molecular oxygen as the precursor gases. Details of the growth and characterization of the resulting films have been reported in Ref. 7. Following deposition the samples were annealed in a rapid thermal annealing furnace under an argon atmosphere at 900°C for 1 minute to densify the oxide.
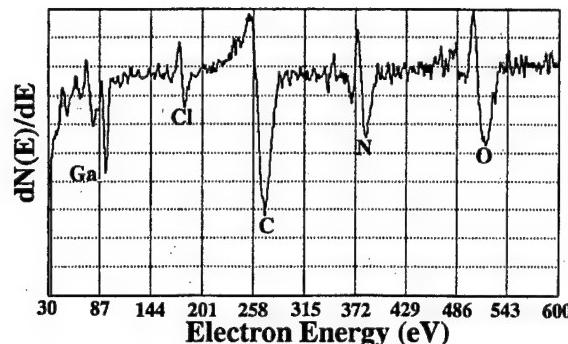
Figure 1.   AES of as-loaded GaN surface after *ex situ* clean described in text.
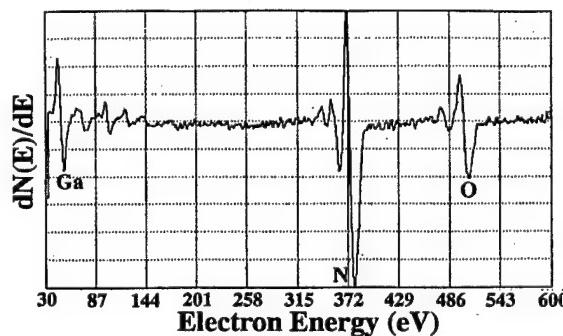
Figure 2.   AES of GaN surface following 30W N_2 plasma clean at 300°C.

Metal insulator semiconductor capacitors were fabricated using all three insulators. These devices were fabricated laterally, i.e., both the Al gate contact and the Al ohmic contact were deposited via thermal evaporation using a shadow mask on the surface of the structure, because of the insulating AlN buffer layer on which of the GaN epilayer was grown. Each of the Ga_2O_3/GaN structures was partially masked with photoresist and reactively ion etched with SF_6 to the GaN layer on which was deposited the ohmic contact. Prior to metallization, each sample was annealed rapid thermally under N_2O at 900°C for 15 seconds. Capacitance-voltage measurements were performed at 1 MHz with an HP 4284A precision LCR meter in conjunction with HP VEE software. Additional capacitors were also fabricated following the same procedure with the exception of annealing in forming gas (10% H_2 in N_2).

The AlN/GaN MIS capacitors were fabricated in a similar manner; however, the ohmic contact was not deposited directly on the GaN. A large area of the AlN was covered with Al, and the gate contacts were deposited on the remainder of the AlN. This follows from the theory that the area of the capacitor created by the large contact creates a capacitance in series with the MIS capacitor which is sufficiently large that it can be considered a short circuit. Capacitance-voltage measurements were performed at 1 MHz both in the dark and following exposure to UV light in an attempt to generate minority carriers.

The deposited SiO_2/GaN MIS capacitors were fabricated in exactly the same manner as the AlN capacitors. Capacitance-voltage measurements were performed at 1 MHz in the dark following the metallization. The capacitors were subsequently annealed in forming gas (10% H_2 in N_2) at 400°C for 30 minutes, and the C-V measurements were repeated.

C. Results and Discussion

Gallium Oxide. No discernible Ga_2O_3 layer was detected by EDS or XPS below 800°C. Films grown at 1000°C had a much rougher surface than those grown at temperatures closer to 800°C, as shown in Fig. 3 and 5. Figures 4 and 6 provide the EDS results which shows that the films contain oxygen and gallium with small nitrogen peaks, the last due to the underlying GaN layer.
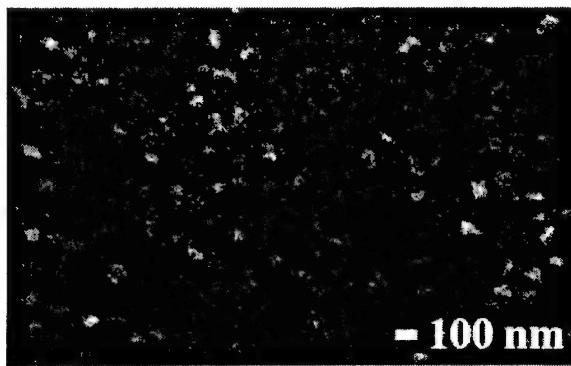

Figure 3.   SEM of Ga_2O_3 grown at 825°C.
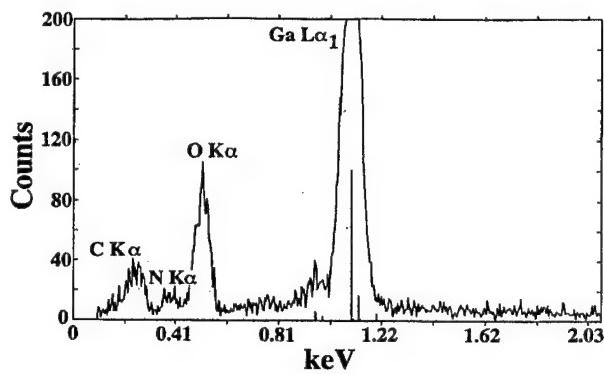
Figure 4.   EDS of Ga$_2$O$_3$ grown at 825°C.
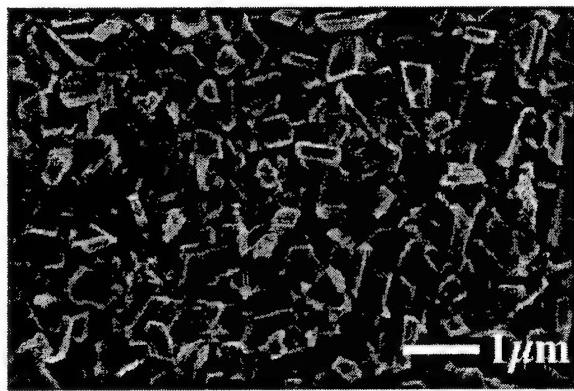
Figure 5.   SEM of Ga$_2$O$_3$ grown at 1000°C.
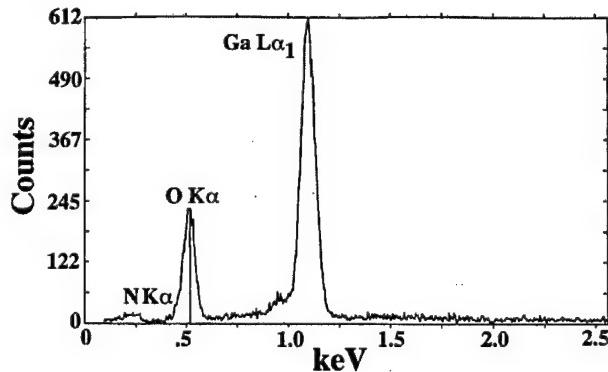

Figure 6.   EDS of Ga$_2$O$_3$ grown at 1000°C.

The MIS capacitors discussed below were fabricated from the oxide grown at 825°C. The resulting C-V demonstrated accumulation and depletion with large leakage. Additional data from samples annealed in N$_2$O and forming gas are presented in Fig. 7 and corrected for 420-470 Ω of series resistance, likely due to the lateral device structure. The C-V measurements did show distinct accumulation and depletion regions; however, neither of the annealed samples showed a reduction in the leakage current.

137
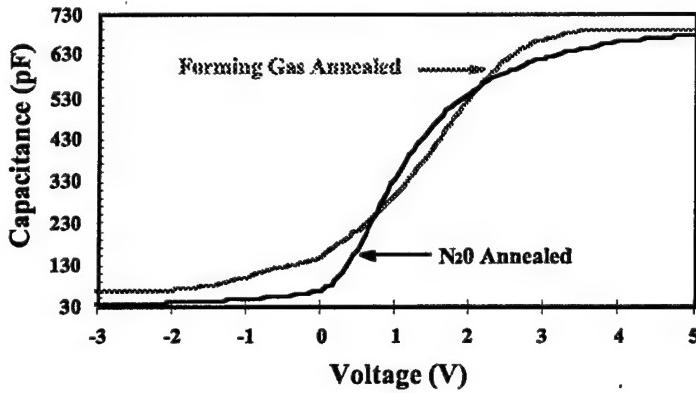
Figure 7.   1 MHz C-V measurement of Ga_2O_3/GaN MIS capacitors.

Aluminum Nitride. The AlN/GaN capacitors were theoretically modeled using the equations presented by Tsividis.[8] The theoretical flatband voltage, V_{FB}, can be determined from the difference in the work-functions of the metal and the semiconductor.

$$V_{FB} = \phi_m - \phi_S \qquad (1)$$

Little information is available concerning the work-function of GaN; however an electron affinity value of 4.1 eV has been reported.[9] Combined with the 4.1 eV value of the work-function of Al, the theoretical value of V_{FB} is zero. The effects of minority carriers are not accounted for in the calculation because in wide bandgap materials the generation rates of minority carriers are extremely slow and lifetimes very short. Casey, *et al.*[2] reported a generation rate of $\approx 4.2 \times 10^{18}$ seconds and a lifetime of 1×10^{-8} seconds. Therefore the minority carriers cannot play a role in the C-V measurements. The C-V measurements, seen in Fig. 8, performed on the AlN/GaN capacitors showed a shift to the left of the flatband voltage indicating a net positive charge in the insulator or at the interface. The curve is also stretched out, possibly due to interface traps located at the interface. A value for the density of interface traps, of $\approx 1 \times 10^{12}$ cm^{-2} eV^{-1} was obtained from the conductance method described in Nicollian and Brews.[10] The hysteresis seen in AlN is different than that is typically seen in devices with "slow" donor traps. The devices were swept from depletion to accumulation. When the devices were swept back to depletion, the curve shifted more to the left indicating more positive charge was present than before. Further research is being conducted to determine the nature of the traps and the charge within the AlN.

A second sweep was done following an exposure to UV light just prior to the measurement. The depletion capacitance was higher than without the exposure to the light, as shown in Fig. 9. When the capacitor was swept back from accumulation to depletion the curve

crossed over the sweep from depletion to accumulation thus indicating the generation of minority carriers within the GaN bulk. Both curves were corrected for a series resistance of 235 Ω.

Silicon Dioxide. The same theoretical model used for the AlN/GaN capacitors was employed for the SiO$_2$/GaN capacitors. Figure 10 shows that the measured values and resulting curve closely agree with the theoretical model. The small amount of stretch-out and hysteresis is believed to be due to interface traps. It is also believed that the gallium oxygen bonding is associated with the hysteresis, since this oxide does not exhibit hysteresis on silicon.[7] The areal density of traps estimated from the hysteresis in the C-V prior to post metallization annealing was 5.6×10^{11} cm^{-2}; following the anneal the density was reduced to 2.4×10^{11} cm^{-2}. The curves were corrected as before for a series resistance of $\approx300\Omega$.

D. Conclusions

The deposition and C-V characteristics of Ga$_2$O$_3$, AlN, and SiO$_2$ on GaN MIS capacitors have been investigated. Depletion and accumulation were achieved in the Ga$_2$O$_3$; however,
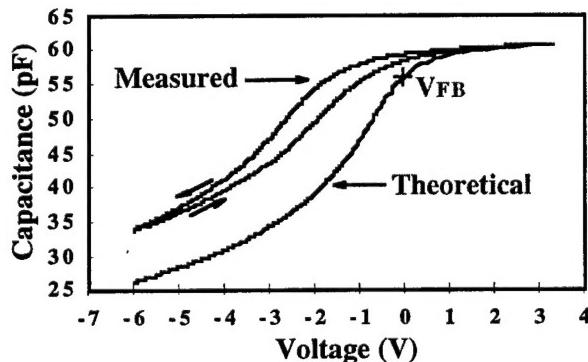
Figure 8.   Comparison of measured and theoretical C-V data for an AlN/GaN capacitor. Measurement conducted at 1MHz in the dark
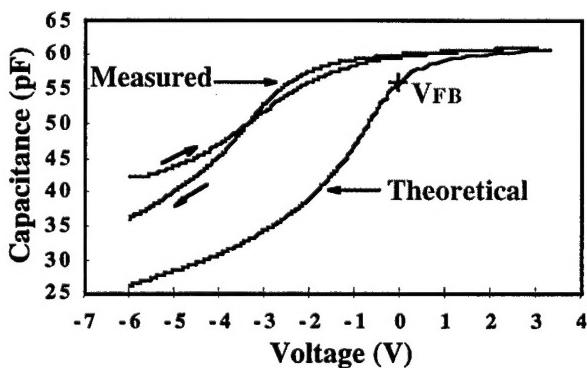
Figure 9.   Comparison of measured and theoretical C-V data for AlN/GaN capacitor. Measurement conducted at 1MHz following 30 sec. UV exposure.
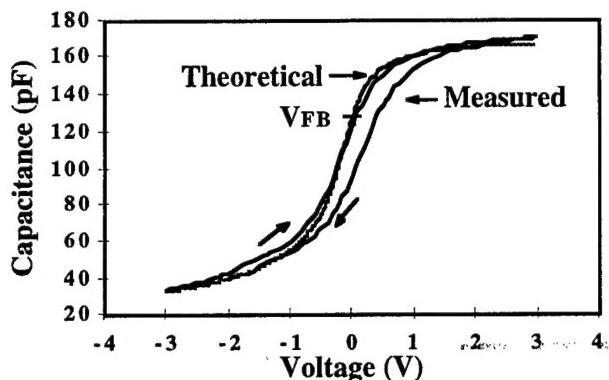
Figure 10.   Theoretical and measured C-V curves of an SiO$_2$/GaN capacitor prior to annealing. Measurement taken at 1 MHz .
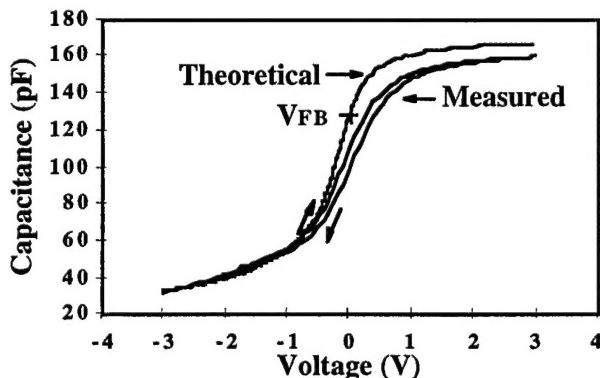
Figure 11.   Theoretical and measured C-V curves of an SiO$_2$/GaN capacitor following an anneal at 400°C in forming gas. Measurement taken at 1 MHz.

better performance was achieved in the AlN and the SiO$_2$. A better understanding of the optimum process route and the characterization of the resulting film must be achieved before being used in MIS structures. The AlN-based capacitors were significantly better than the Ga$_2$O$_3$ because of the lower leakage; however, the films are not of device quality. The large amount of hysteresis and charge in the structure must be eliminated. The SiO$_2$ films showed the most promise as an insulator in GaN FETs. The SiO$_2$ also has many advantages over the AlN as a result of the lower deposition temperatures and ease of processing. The electrical characteristics of the SiO$_2$-based capacitors were superior to the other insulators studied in this research. Additional investigations regarding surface preparation of the GaN and the deposition conditions for the SiO$_2$ should provide both an improved insulator and effective utilization of the unique combination of physical and electronic properties of GaN in MISFET devices.

E. Future Work

Studies of the AlN/GaN and the SiO$_2$/GaN MIS structures will continue along with studies of Si3N4/GaN structures. MISFET devices utilizing the above insulating materials will be fabricated and characterized.

F. Acknowledgments

G. References

1. M. Yoder, IEEE Trans. Elect. Dev. **43** (10), 1633 (1996).
2. H.C. Casey, Jr., G.G. Fountain, R.G. Alley, B.P. Keller, Steven P. Denbarrs, Appl. Phys. Lett. **68** (13), 1850 (1996).
3. S.C Binari, Presentation at ISCS Conference, Sept 1994.
4. S.D. Wolter, B.P. Luther, D.L. Waltemyer, C. Onneby, S.E. Mohney, R.J. Molnar, Appl. Phys. Lett. **70** (16), 2156 (1997).
5. T.W. Weeks, Jr., M.D. Bremser, K.S. Ailey, E. Carlson, W.G. Perry, E.L. Piner, N.A. El-Masery, R.F. Davis, J. Mater. Res. **11** (4), 1011 (1996.).
6. S.W.King, Ph.D. Thesis, North Carolina State University, 1997
7. T. Yasuda, Y. Ma, S. Habermehl, G. Lucovsky, Appl. Phys. Lett. **60** (4), 434 (1992).
8. Y.P. Tsividis, *Operation and Modeling of the MOS Transistor*, (McGraw-Hill, Inc., New York, 1987), p. 476.
9. J.I. Pankove and H. Schade, Appl. Phys. Lett. **25**, 53 (1974).
10. E. H. Nicollian and J.R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*, (Wiley, New York, 1982), p. 214.

XIII. Distribution List

Dr. Colin Wood   3
Office of Naval Research
Electronics Division, Code: 312
Ballston Tower One
800 N. Quincy Street
Arlington, VA 22217-5660

Administrative Contracting Officer   1
Office of Naval Research
Regional Office Atlanta
100 Alabama Street, Suite 4R15
Atlanta, GA 30303

Director, Naval Research Laboratory   1
ATTN: Code 2627
Washington, DC 20375

Defense Technical Information Center   2
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218